Linear Algebra I (Math 3260)

Sean F. Ellermeyer & Lake R. Ritter

Department of Mathematics
Kennesaw State University

How to use this text: This text is meant to be read actively (i.e., with pen and paper in hand). Exercises, labeled in bold, are strategically embedded within the narrative of this text, and it is recommended that you attempt each exercise as you encounter it. If, after an honest attempt, you can't complete an exercise, move on with a plan to come back. Your questions are a welcomed contribution to class sessions!

Reporting Errata: To report errors, please email the authors at

- sellerme@kennesaw.edu, or
- lritter@kennesaw.edu.

Contents

Chapter 1

The Vector Spaces R^n

We will begin our study of linear algebra by introducing the vector spaces R^n, which are also known as the Euclidean n–spaces. As a student in an introductory linear algebra course, you have probably completed Calculus I and you are probably comfortable with working in the mathematical settings of R and R^2. You have learned to visualize R as the *number line*. It is a one–dimensional object – a line that has infinite length. You have also learned to visualize R^2 as the *Cartesian plane*. It is a two–dimensional object – an infinite flat plane. Perhaps you have also completed (or are concurrently taking) Calculus III and, if so, you have learned to visualize R^3 as *three–dimensional space*. As human beings, we have the ability to form visual pictures of R, R^2, and R^3 because the physical world that we live in is three–dimensional. When it comes to trying to visualize R^4 (or R^n for any $n \geq 4$), we don't have the ability to form visual pictures. However, as will be seen, we can still equip ourselves with the mathematical tools that are needed to address problems involving R^n when $n \geq 4$.

1.1 The Vector Space R^2

1.1.1 What is a Vector in R^2?

If $P = (p_1, p_2)$ and $Q = (q_1, q_2)$ are two points in R^2, then the directed line segment from P to Q is the arrow that points from P to Q. We denote this directed line segment by $\overrightarrow{PQ}$. It is called a *directed* line segment because it is thought of as "beginning" at the point P and "ending" at the point Q. As an example, the directed line segment from the point $P = (2, 3)$ to the point $Q = (5, 10)$ is pictured in Figure 1.1.

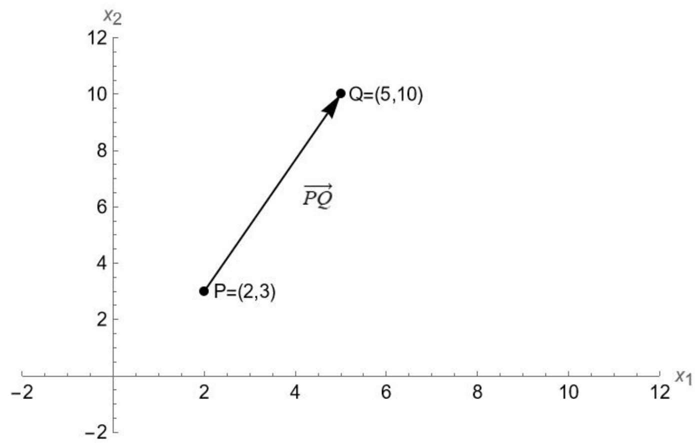


Figure 1.1: Directed Line Segment $\overrightarrow{PQ}$

Two directed line segments are said to be equivalent to each other if they both have the same length and both point in the same direction. For example the directed line segment $\overrightarrow{PQ}$ from the point $P = (2, 3)$ to the point $Q = (5, 10)$ is equivalent to the directed line segment $\overrightarrow{RS}$ from the point $R = (6, 1)$ to the point $S = (9, 8)$ because $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ both have the same length and both point in the same direction. The reason that $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ both have the same length and both point in the same direction is that the ending point of each of these directed line segments is reached from the starting point by moving 3 units to the right and 7 units upward, as illustrated in Figure 1.2.

Now that we have described what a directed line segment is and what it means for two directed line segments to be equivalent, we can describe what we mean by a vector: A **vector** in R^2 is an ordered pair of real numbers that describes both a length (also called a "magnitude") and a direction. We use the triangular bracket notation $\langle x_1, x_2 \rangle$ for vectors, so as not to confuse vectors with points. (The rounded bracket notation (x_1, x_2) is used for points.) When we want to give a name to a vector, we use a notation such as $\vec{x}$ or $\mathbf{x}$. Thus we could write $\vec{x} = \langle x_1, x_2 \rangle$ or $\mathbf{x} = \langle x_1, x_2 \rangle$. The real numbers, x_1 and x_2, are called the **entries** or the **components** of the vector $\vec{x} = \langle x_1, x_2 \rangle$. If $\vec{x} = \langle x_1, x_2 \rangle$ is a vector in R^2, then we can visualize $\vec{x}$ by drawing its

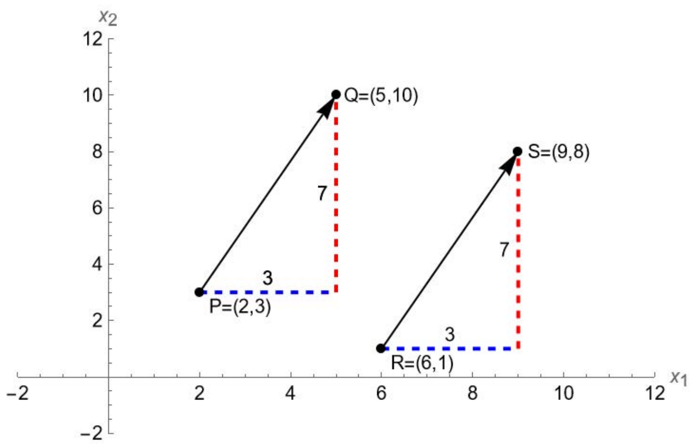


Figure 1.2: Equivalence of $\overrightarrow{PQ}$ and $\overrightarrow{RS}$

standard representative, which is the directed line segment from the point $O = (0, 0)$ to the point $X = (x_1, x_2)$. We can also visualize $\vec{x} = \langle x_1, x_2\rangle$ by drawing any other representative of $\vec{x}$. This is done by choosing any point $P = (p_1, p_2)$ and then drawing the directed line segment from P to the point $Q = (p_1 + x_1, p_2 + x_2)$.

To illustrate: The standard representative of the vector $\vec{x} = \langle 3, 7\rangle$ is the directed line segment $\overrightarrow{OX}$ that is pictured in Figure 1.3. The directed line segments $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ that are pictured in Figure 1.2 are also representatives of the vector $\vec{x} = \langle 3, 7\rangle$.

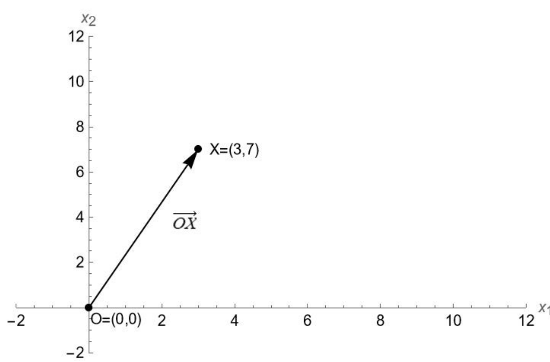


Figure 1.3: Standard Representative of Vector $\vec{x} = \langle 3, 7\rangle$

Exercise 1.1.1. *Draw a picture of the standard representative of the vector $\vec{x} = \langle -3, 4 \rangle$. Then draw a picture of the representative of $\vec{x}$ that is based at the point $P = (1, 2)$. (To do this you will need to find the point Q such that $\overrightarrow{PQ}$ is a representative of $\vec{x}$.)*

Exercise 1.1.2. *What vector is represented by the directed line segment $\overrightarrow{PQ}$ from the point $P = (3, 1)$ to the point $Q = (-4, -1)$? Draw a picture of the standard representative of this vector.*

Exercise 1.1.3. *In parts 1–5 below, four points (P, Q, R, and S) are given. Draw the directed line segments $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ and determine whether or not $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ represent the same vector. If they do not represent the same vector, then state whether this is because they don't have the same length or don't point in the same direction (or both).*

1. $P = (-7, -7)$, $Q = (4, -3)$, $R = (-1, 7)$, $S = (10, 11)$
2. $P = (1, -6)$, $Q = (-7, -5)$, $R = (-6, -8)$, $S = (-14, -7)$
3. $P = (-2, 6)$, $Q = (-8, 7)$, $R = (-8, -5)$, $S = (-1, -5)$
4. $P = (-8, 0)$, $Q = (5, 6)$, $R = (3, 3)$, $S = (-10, -3)$
5. $P = (4, 7)$, $Q = (-8, -8)$, $R = (0, 1)$, $S = (-4, -4)$

1.1.2 Addition of Vectors

Having defined what is meant by a vector in R^2, we will define two operations on vectors. One operation is called vector addition. It is used to add two vectors in R^2 to obtain another vector in R^2. The other operation is called scalar multiplication. It is used to multiply a vector in R^2 by a real number (referred to as a scalar) in order to obtain another vector in R^2. We will first describe vector addition.

If $\vec{x} = \langle x_1, x_2 \rangle$ and $\vec{y} = \langle y_1, y_2 \rangle$ are two vectors in R^2, then we define the **vector sum** $\vec{x} + \vec{y}$ to be the vector

$$\vec{x} + \vec{y} = \langle x_1 + y_1, x_2 + y_2 \rangle .$$

It is easy to compute the sum of two given vectors. For example, suppose that $\vec{x} = \langle 4, -4 \rangle$ and $\vec{y} = \langle 5, 0 \rangle$. Then the vector sum of $\vec{x}$ and $\vec{y}$ is

$$\vec{x} + \vec{y} = \langle 4 + 5, -4 + 0 \rangle = \langle 9, -4 \rangle .$$

Although the calculations that are done to perform vector addition are straightforward, it is important to understand what is really going on with these operations by drawing some pictures. Let us first consider the example

$$\begin{aligned} \vec{x} &= \langle 4, -4 \rangle \\ \vec{y} &= \langle 5, 0 \rangle \\ \vec{x} + \vec{y} &= \langle 9, -4 \rangle \ . \end{aligned}$$

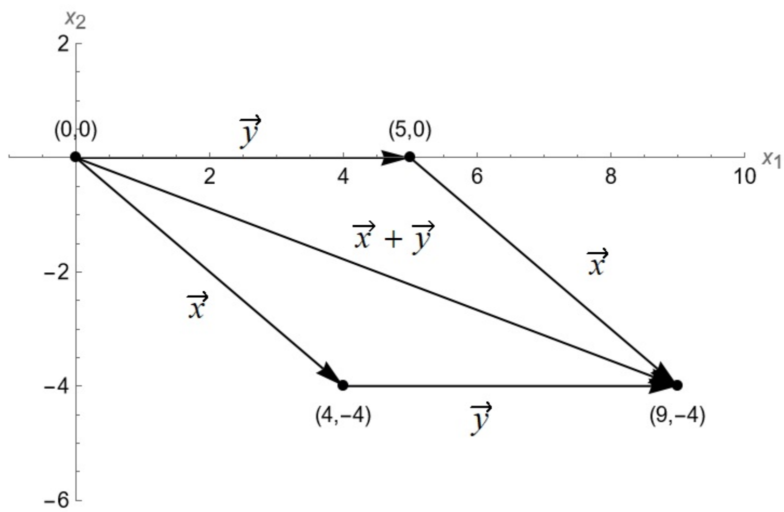


Figure 1.4: Picture of $\vec{x} + \vec{y} = \vec{y} + \vec{x}$

Figure 1.4 shows the standard representatives of the vectors $\vec{x} = \langle 4, -4 \rangle$, $\vec{y} = \langle 5, 0 \rangle$, and $\vec{x} + \vec{y} = \langle 9, -4 \rangle$. Also pictured are the representative of $\vec{x}$ based at the point $(5, 0)$ and the representative of $\vec{y}$ based at the point $(4, -4)$. These four directed line segments form a parallelogram with sides formed by the vectors $\vec{x}$ and $\vec{y}$. The diagonal of the parallelogram is the vector $\vec{x}+\vec{y}$. The idea here is that to draw $\vec{x}+\vec{y}$, we start at the point $(0, 0)$, then travel along the vector $\vec{x}$ to the point $(4, -4)$, and then travel from $(4, -4)$ along the vector $\vec{y}$ to arrive at the point $(9, -4)$. On the other hand, to draw $\vec{y} + \vec{x}$, we start at $(0, 0)$, then travel along the vector $\vec{y}$ to the point $(5, 0)$, and then travel from $(5, 0)$ along the vector $\vec{x}$ to arrive at the point $(9, -4)$. So $\vec{x}+\vec{y}$ is the same as $\vec{y}+\vec{x}$. It is easily seen by direct computation that $\vec{x} + \vec{y}$ is the same as $\vec{y} + \vec{x}$, but the parallelogram picture is a helpful visual aid in seeing why this is true geometrically.

The above example illustrates the general procedure for drawing a picture of the vector sum of any two given vectors $\vec{x}$ and $\vec{y}$. This procedure, is called the *Parallelogram Method of Vector Addition.*

The Parallelogram Method of Vector Addition
(Refer to Figure 1.5.)

To illustrate the vector sum of vectors $\vec{x}$ and $\vec{y}$:

1. Draw pictures of the standard representative $\overrightarrow{OX}$ of $\vec{x}$, and the standard representative $\overrightarrow{OY}$ of $\vec{y}$.

2. Draw the representative of $\vec{x}$ based at the point Y and draw the representative of $\vec{y}$ based at the point X. These two representatives will both end at a common point R.

3. After completing the first two steps, you should have a picture of a parallelogram, unless the vectors $\vec{x}$ and $\vec{y}$ point in the same or opposite directions, in which case you will just have a picture of a line segment. In either case, the directed line segment $\overrightarrow{OR}$ is the standard representative of the vector $\vec{x} + \vec{y}$.

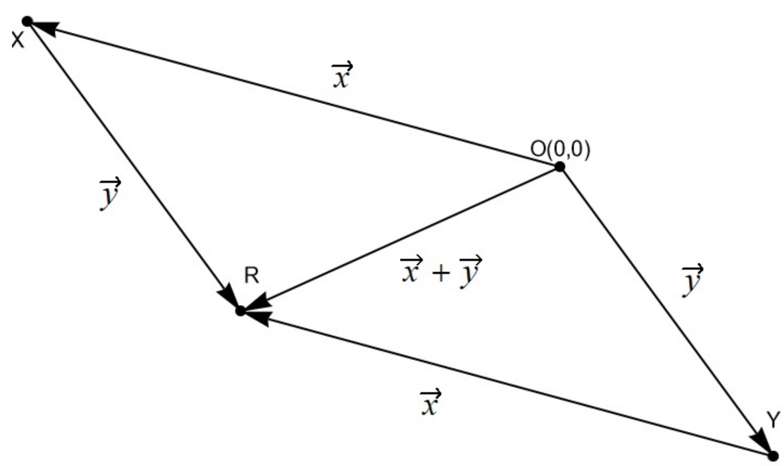


Figure 1.5: The Parallelogram Method of Vector Addition

Exercise 1.1.4. *For each pair of vectors, $\vec{x}$ and $\vec{y}$, given in parts 1–7, compute $\vec{x}+\vec{y}$ and then draw a picture to illustrate the Parallelogram Method of Vector Addition for $\vec{x}+\vec{y}$.*

1. $\vec{x} = \langle 3, -4 \rangle$, $\vec{y} = \langle 4, -2 \rangle$
2. $\vec{x} = \langle 0, 1 \rangle$, $\vec{y} = \langle -3, -4 \rangle$
3. $\vec{x} = \langle 3, 0 \rangle$, $\vec{y} = \langle 0, -1 \rangle$
4. $\vec{x} = \langle -3, 3 \rangle$, $\vec{y} = \langle 0, 3 \rangle$
5. $\vec{x} = \langle -3, 3 \rangle$, $\vec{y} = \langle -3, 3 \rangle$
6. $\vec{x} = \langle -3, 3 \rangle$, $\vec{y} = \langle 6, -6 \rangle$
7. $\vec{x} = \langle -3, 3 \rangle$, $\vec{y} = \langle 3, -3 \rangle$

1.1.3 The Zero Vector and Additive Inverses

The **zero vector** in R^2 is the vector $\langle 0, 0 \rangle$. It differs from any other vector in R^2 in that it is the only vector that does not have any length and does not point in any direction. It is represented by a point rather than by a directed line segment. Whereas any other vector in R^2 is a one–dimensional object (because it has a dimension of length), the zero vector is a zero–dimensional object (because it has no length). We will use the notation $\vec{0}_2 = \langle 0, 0 \rangle$ to denote the zero vector in R^2.

Our first observation about the zero vector is that it is the **additive identity** vector in R^2. What this means is that if $\vec{x}$ is any vector in R^2, then $\vec{x}+\vec{0}_2 = \vec{x}$. This is easily seen to be true because if $\vec{x} = \langle x_1, x_2 \rangle$ is any vector in R^2 then

$$\vec{x} + \vec{0}_2 = \langle x_1, x_2 \rangle + \langle 0, 0 \rangle = \langle x_1 + 0, x_2 + 0 \rangle = \langle x_1, x_2 \rangle = \vec{x}.$$

Our second observation about the zero vector is that if $\vec{x}$ is any vector in R^2, then there is some other vector in R^2, called the **additive inverse** of $\vec{x}$ and denoted by $-\vec{x}$, such that $\vec{x} + (-\vec{x}) = \vec{0}_2$. If we are given a particular vector $\vec{x} = \langle x_1, x_2 \rangle$, it is easily seen that the additive inverse of $\vec{x}$ is $-\vec{x} = \langle -x_1, -x_2 \rangle$ because

$$\langle x_1, x_2 \rangle + \langle -x_1, -x_2 \rangle = \langle x_1 - x_1, x_2 - x_2 \rangle = \langle 0, 0 \rangle .$$

For example the additive inverse of the vector $\vec{x} = \langle -4, 1 \rangle$ is $-\vec{x} = \langle 4, -1 \rangle$ because $\langle -4, 1 \rangle + \langle 4, -1 \rangle = \langle 0, 0 \rangle$.

Exercise 1.1.5. *What is the additive inverse of the vector $\langle -4, 4 \rangle$? What is the additive inverse of the vector $\langle 3, 0 \rangle$? What is the additive inverse of the vector $\langle 0, 0 \rangle$?*

Having defined vector addition and also having defined what we mean by the additive inverse of a vector, we can now define vector subtraction. If $\vec{x}$ and $\vec{y}$ are two vectors, then we define the **vector difference** $\vec{x} - \vec{y}$ to be the sum of $\vec{x}$ and the additive inverse of $\vec{y}$. In other words,

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

So, for example if $\vec{x} = \langle -4, 1 \rangle$ and $\vec{y} = \langle 3, 5 \rangle$, then

$$\vec{x} - \vec{y} = \vec{x} + (-\vec{y}) = \langle -4, 1 \rangle + \langle -3, -5 \rangle = \langle -7, -4 \rangle .$$

Exercise 1.1.6. *For the vectors $\vec{x} = \langle -5, -1 \rangle$ and $\vec{y} = \langle 6, 3 \rangle$, compute $\vec{x} - \vec{y}$ and $\vec{y} - \vec{x}$.*

1.1.4 Scalar Multiples of Vectors

If $\vec{x} = \langle x_1, x_2 \rangle$ is a vector in R^2 and c is a real number (also called a scalar), then we define the **scalar multiple** of the vector $\vec{x}$ by the scalar c to be the vector

$$c\vec{x} = \langle cx_1, cx_2 \rangle .$$

For example if $\vec{x} = \langle 2, -3 \rangle$ and $c = 3$, then the scalar multiple of $\vec{x}$ by c is

$$3\vec{x} = \langle 3\,(2)\,, 3\,(-3) \rangle = \langle 6, -9 \rangle .$$

Just as drawing pictures helps us to understand the concept of vector addition, it also helps us to understand the concept of scalar multiplication. For example, if $\vec{x} = \langle 2, -3 \rangle$, then

$$3\vec{x} = \langle 6, -9 \rangle = \langle 2, -3 \rangle + \langle 2, -3 \rangle + \langle 2, -3 \rangle = \vec{x} + \vec{x} + \vec{x}$$

and it can be seen (as illustrated in Figure 1.6), that the vector $3\vec{x}$ points in the same direction as the vector $\vec{x}$ and has 3 times the length of the vector $\vec{x}$.

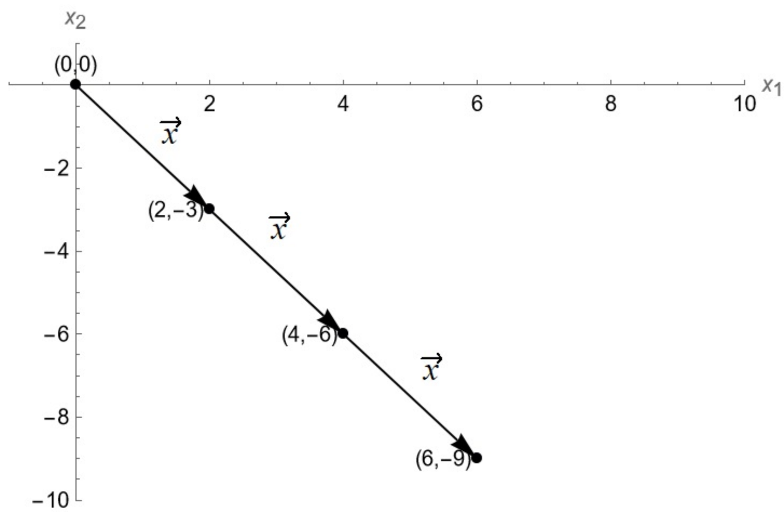


Figure 1.6: $3\vec{x} = \vec{x} + \vec{x} + \vec{x}$

As another example, if $\vec{x} = \langle 2, -3 \rangle$, then $-2\vec{x} = \langle -4, 6 \rangle$ and it can be seen in Figure 1.7 that the vector $-2\vec{x}$ points in the opposite direction of the vector $\vec{x}$ and has 2 times the length of $\vec{x}$.

In general, if $\vec{x}$ is any nonzero vector in R^2 and c is any nonzero scalar, then

- $c\vec{x}$ points in the direction of $\vec{x}$ when $c > 0$ and in the opposite direction (i.e., $180°$) of $\vec{x}$ when $c < 0$.
- the length of $c\vec{x}$ is $|c|$ times the length of $\vec{x}$.

We can summarize the cases involving the zero vector or the zero scalar. In particular, if c is any scalar, then

$$c\vec{0}_2 = c\,\langle 0, 0 \rangle = \langle 0, 0 \rangle = \vec{0}_2.$$

Likewise, if $\vec{x} = \langle x_1, x_2 \rangle$ is any vector, then

$$0\vec{x} = 0\,\langle x_1, x_2 \rangle = \langle 0, 0 \rangle = \vec{0}_2.$$

Exercise 1.1.7. *For the vector* $\vec{x} = \langle 2, 4 \rangle$*:*

1. *Compute* $3\vec{x}$ *and draw a picture that shows both* $\vec{x}$ *and* $3\vec{x}$*.*

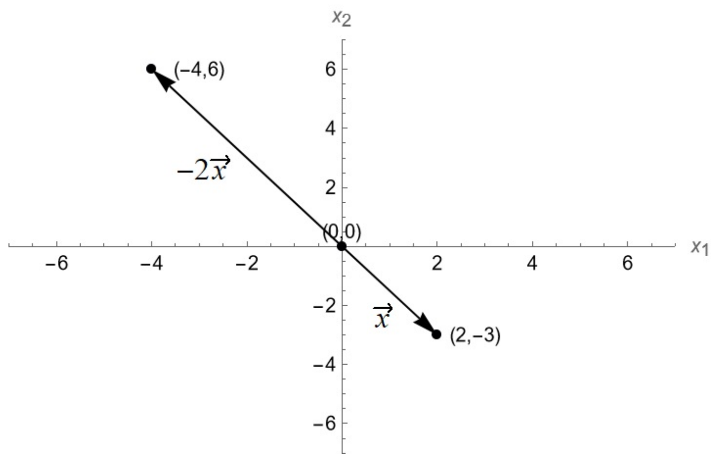


Figure 1.7: Picture of $\vec{x}$ and $-2\vec{x}$

2. *Compute $-2\vec{x}$ and draw a picture that shows both $\vec{x}$ and $-2\vec{x}$.*

3. *Compute $0\vec{x}$ and draw a picture that shows both $\vec{x}$ and $0\vec{x}$.*

Exercise 1.1.8. *Fill in the blanks to correctly complete the sentences below.*

1. *If $\vec{x}$ is any vector in R^2 with $\vec{x} \neq \vec{0}_2$, then the vector $2\vec{x}$ points in the _______ direction of $\vec{x}$ and has ______ times the length of $\vec{x}$.*

2. *If $\vec{x}$ is any vector in R^2 with $\vec{x} \neq \vec{0}_2$, then the vector $\frac{1}{3}\vec{x}$ points in the _______ direction of $\vec{x}$ and has ______ times the length of $\vec{x}$.*

3. *If $\vec{x}$ is any vector in R^2 with $\vec{x} \neq \vec{0}_2$, then the vector $-3\vec{x}$ points in the _______ direction of $\vec{x}$ and has ______ times the length of $\vec{x}$.*

4. *If $\vec{x}$ is any vector in R^2 with $\vec{x} \neq \vec{0}_2$, then the vector $-\frac{1}{5}\vec{x}$ points in the _______ direction of $\vec{x}$ and has ______ times the length of $\vec{x}$.*

5. *If $\vec{x}$ is any vector in R^2, then $0\vec{x} =$ ______.*

6. *If c is any scalar, then $c\vec{0}_2 =$ _____.*

Exercise 1.1.9. *For each of the vector pairs $\vec{x}$ and $\vec{y}$ and each of the scalar pairs c and d given below, compute $c\vec{x} + d\vec{y}$. Draw a picture that contains $c\vec{x}$, $d\vec{y}$ and $c\vec{x} + d\vec{y}$.*

1. $\vec{x} = \langle -4, -4\rangle$, $\vec{y} = \langle -3, -2\rangle$, $c = 1$, $d = 4$.

2. $\vec{x} = \langle -4, -2\rangle$, $\vec{y} = \langle -4, 4\rangle$, $c = 3$, $d = 5$.

3. $\vec{x} = \langle 0, 2\rangle$, $\vec{y} = \langle -2, -5\rangle$, $c = -5$, $d = -1$.

4. $\vec{x} = \langle -4, 5\rangle$, $\vec{y} = \langle 0, 3\rangle$, $c = 2$, $d = -5$.

1.1.5 Linear Combinations in R^2

Given a pair of vectors $\vec{x}$ and $\vec{y}$ in R^2 and a pair of scalars, c and d, we have a special name for a vector of the form $c\vec{x} + d\vec{y}$, like those appearing in Exercise 1.1.9. We refer to $c\vec{x} + d\vec{y}$ as a *linear combination* of the vectors $\vec{x}$ and $\vec{y}$. More generally, whenever we use these two key operations that we've defined, vector addition and scalar multiplication, on any collection of vectors (not just two) we call the result a **linear combination**. We can even apply the phrase when dealing with a single vector. That is, given a vector $\vec{x}$, and a scalar c, the vector $c\vec{x}$ is a linear combination of the vector $\vec{x}$. On occasion, we will be interested in the collection of all linear combinations of a set of vectors. In such cases, we allow the scalars to vary over all real numbers.

Example 1.1.1. *Let $\vec{e}_1 = \langle 1, 0\rangle$. Give a geometric description of the set of all linear combinations of the vector $\vec{e}_1$.*

We note that if we plot vector $\vec{e}_1$ as a directed line segment from the point $(0, 0)$ to the point $(1, 0)$, we get a horizontal line segment. If $\vec{x}$ is any linear combination of $\vec{e}_1$, then

$$\vec{x} = c\vec{e}_1 = \langle c(1), c(0)\rangle = \langle c, 0\rangle.$$

If $c = 0$, we get the zero vector. For $c \neq 0$ we get a horizontal vector that we could plot as a directed line segment starting at the origin and ending at the point $(c, 0)$ on the horizontal (a.k.a. the x)-axis. If $c < 0$, our line segment would terminate at some point on the negative x-axis, and if $c > 0$, our line segment would terminate at some point on the positive x-axis. Hence, if we allow c to vary over all real numbers, we could associate the set of all linear combination of $\vec{e}_1$ with the whole x-axis. Thus, we can say that the set of all linear combinations of $\vec{e}_1$ is the x-axis in our Cartesian plane.

Exercise 1.1.10. *Using the reasoning demonstrated in Example 1.1.1, give a geometric description of the collection of all linear combinations of each vector.*

1. $\vec{e}_2 = \langle 0, 1 \rangle$
2. $\vec{u} = \langle 1, 1 \rangle$

1.1.6 Magnitude, Dot Product, and Orthogonality

If $\vec{x} = \langle x_1, x_2 \rangle$ is a vector in R^2, then the length of $\vec{x}$ can be determined by using the Pythagorean Theorem. If we draw the standard representative of $\vec{x}$ (or any representative of $\vec{x}$), we can see that $\vec{x}$ is the hypotenuse of a right triangle with side lengths $|x_1|$ and $|x_2|$. (We have put absolute values on x_1 and x_2 in case either one of these numbers is negative, since lengths cannot be negative.) This is illustrated in Figure 1.8.

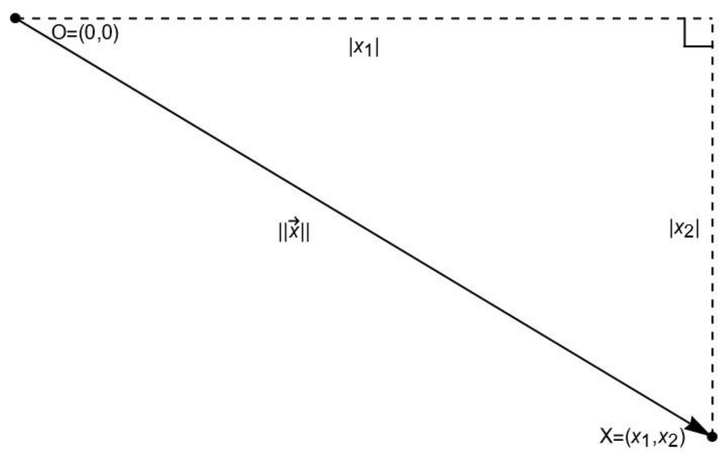


Figure 1.8: $\|\vec{x}\|^2 = x_1^2 + x_2^2$

Using the notation $\|\vec{x}\|$ to denote the length of $\vec{x}$, the Pythagorean Theorem tells us that

$$\|\vec{x}\|^2 = |x_1|^2 + |x_2|^2 .$$

Since $|x_1|^2 = x_1^2$ and $|x_2|^2 = x_2^2$, we can drop the absolute values in the above equation and just write the equation as

$$\|\vec{x}\|^2 = x_1^2 + x_2^2.$$

Taking the square root of both sides of the above equation gives

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

but since $\|\vec{x}\|$ is a length, it cannot be negative and we thus have

$$\text{Length of } \vec{x} = \|\vec{x}\| = \sqrt{x_1^2 + x_2^2}.$$

The length of $\vec{x}$ is also called the **magnitude** of $\vec{x}$.

Definition 1.1.1. *The* ***magnitude*** *(also called* ***length****) of a vector* $\vec{x} = \langle x_1, x_2 \rangle$ *in* R^2 *is denoted by* $\|\vec{x}\|$ *and is defined to be*

$$\|\vec{x}\| = \sqrt{x_1^2 + x_2^2}.$$

As an example, suppose that $\vec{x} = \langle 1, 5 \rangle$. Then the length of $\vec{x}$ is

$$\|\vec{x}\| = \sqrt{1^2 + 5^2} = \sqrt{26} \approx 5.099.$$

As another example, the length of $\vec{x} = \langle -3, 1 \rangle$ is

$$\|\vec{x}\| = \sqrt{(-3)^2 + 1^2} = \sqrt{10} \approx 3.162.$$

We have a special name for a vector having length one. We call such a vector a **unit vector**. (In some settings, such as physics, it's even customary to use different notation, such as $\hat{u}$ instead of $\vec{u}$ to indicate that a vector is a unit vector.) Unit vectors can be useful for applications in which the direction of a vector is critical but the magnitude is of little interest. For $c \neq 0$, a nonzero vector $\vec{x}$ and its scalar multiple $c\vec{x}$ are parallel—they have the same direction or opposite (180°) directions. Since scalar multiplication affects magnitude, $\|c\vec{x}\| = |c|\|\vec{x}\|$, given any nonzero vector $\vec{x}$ we can obtain a unit vector parallel to $\vec{x}$ (see Exercise 1.1.14).

Exercise 1.1.11. *Draw pictures of each of the following vectors and compute their lengths.*

1. $\vec{x} = \langle -4, 3 \rangle$
2. $\vec{x} = \langle -3, 4 \rangle$

3. $\vec{x} = \langle -6, 4 \rangle$

4. $\vec{x} = \langle 1, 0 \rangle$

5. $\vec{x} = \langle 0, 0 \rangle$.

Exercise 1.1.12.

1. *Explain why, if $\vec{x}$ is any vector in R^2, the additive inverse of $\vec{x}$ has the same length as $\vec{x}$. In other words, explain why*

$$\|-\vec{x}\| = \|\vec{x}\| .$$

2. *Explain why, if $\vec{x}$ and $\vec{y}$ are any two vector in R^2, the vectors $\vec{x} - \vec{y}$ and $\vec{y} - \vec{x}$ have the same length. In other words, explain why*

$$\|\vec{x} - \vec{y}\| = \|\vec{y} - \vec{x}\| .$$

Exercise 1.1.13.

1. *Let $\vec{x} = \langle -3, 4 \rangle$. Compute the lengths of $\vec{x}$ and $2\vec{x}$.*

2. *Let $\vec{x} = \langle -3, 4 \rangle$. Compute the lengths of $\vec{x}$ and $-3\vec{x}$.*

3. *Explain why, if $\vec{x}$ is any vector in R^2 and c is any scalar, then the length of $c\vec{x}$ is equal to the absolute value of c times the length of $\vec{x}$. In other words, explain why*

$$\|c\vec{x}\| = |c| \, \|\vec{x}\| .$$

Exercise 1.1.14.

1. *Show that $\langle 1, 0 \rangle$ and $\langle 0, 1 \rangle$ are unit vectors.*

2. *Determine whether each of the following vectors is a unit vector.*

   a. $\left\langle \frac{3}{5}, -\frac{4}{5} \right\rangle$
   b. $\langle 1, -1 \rangle$
   c. $\left\langle \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\rangle$
   d. $\left\langle -\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}} \right\rangle$

3. *Show that if* θ *is any angle, the vector* $\langle \cos\theta, \sin\theta\rangle$ *is a unit vector.*

4. *Consider* $\vec{x} = \langle 6, 8\rangle$*. Find a positive number* c *such that* $c\vec{x}$ *is a unit vector.*

5. *Suppose* $\vec{x} = \langle x_1, x_2\rangle$ *is any nonzero vector. Find a positive number* c *such that* $c\vec{x}$ *is a unit vector.*

Next, we will derive an algebraic criterion for determining whether or not two vectors in R^2 are perpendicular to each other. This criterion will be seen to involve the "dot product" of the two vectors.

We say that two non–zero vectors, $\vec{x}$ and $\vec{y}$, in R^2 are **perpendicular** to each other if their standard representatives (or any representatives based at the same point) form a 90° angle. Figure 1.9 shows the standard representatives of the vectors $\vec{x} = \langle -2, 4\rangle$ and $\vec{y} = \langle 2, 0\rangle$. These vectors are not perpendicular to each other, because they do not form a 90° angle. Figure 1.10 shows the standard representatives of the vectors $\vec{z} = \langle -2, 2\rangle$ and $\vec{w} = \langle -1, -1\rangle$, which are perpendicular to each other because they do form a 90° angle.

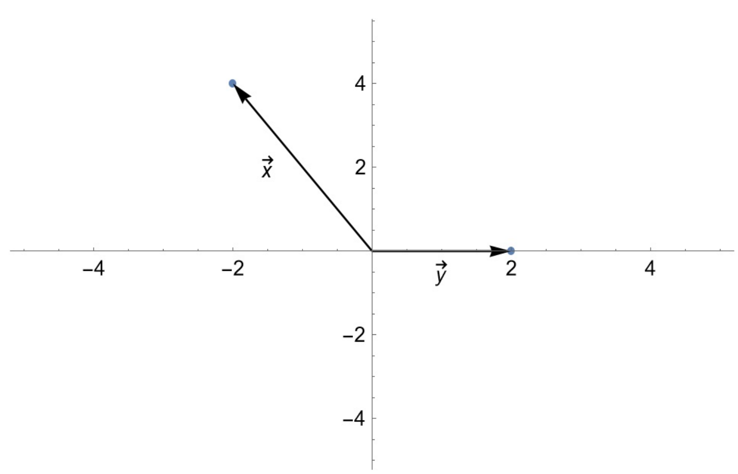


Figure 1.9: $\vec{x}$ and $\vec{y}$ are not perpendicular.

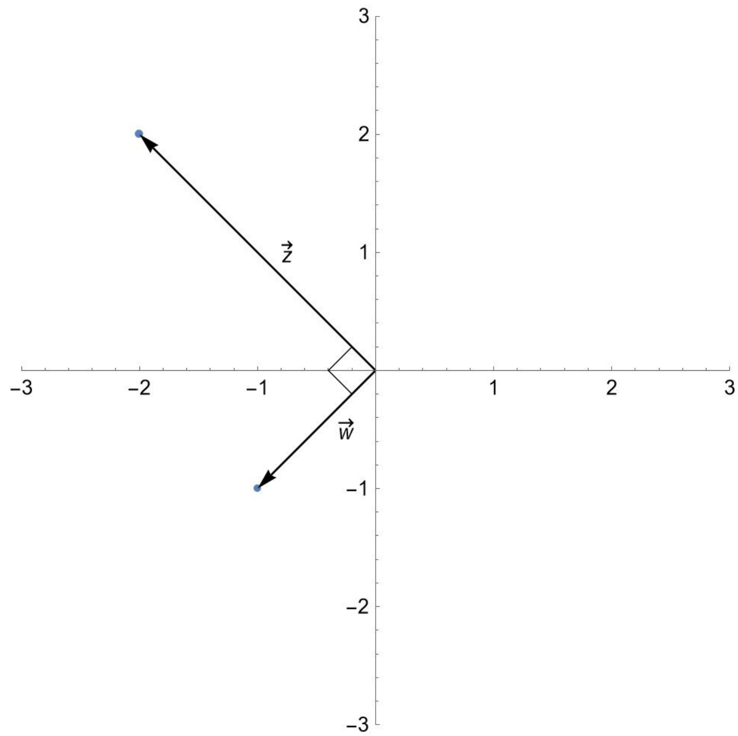


Figure 1.10: $\vec{z}$ and $\vec{w}$ are perpendicular.

The key observation that we need to obtain an algebraic criterion for perpendicularity is that two vectors, $\vec{x}$ and $\vec{y}$, are perpendicular to each other if and only if $\|\vec{x}+\vec{y}\| = \|\vec{x}-\vec{y}\|$. The reason for this can be seen by examining Figures 1.11, 1.12 and 1.13. Recall that we can illustrate the addition of two vectors $\vec{x}$ and $\vec{y}$ by using the parallelogram method. We draw two copies of $\vec{x}$ and two copies of $\vec{y}$ to form a parallelogram. The diagonals of this parallelogram are the vectors $\vec{x}+\vec{y}$ and $\vec{x}-\vec{y}$, as seen in each of Figures 1.11, 1.12, and 1.13. In Figure 1.11, the vectors $\vec{x}$ and $\vec{y}$ form a right ($= 90°$) angle and the parallelogram is actually a rectangle. The two diagonals of the rectangle have equal length, meaning that $\|\vec{x}+\vec{y}\| = \|\vec{x}-\vec{y}\|$. In Figure 1.12, the vectors $\vec{x}$ and $\vec{y}$ form an acute ($< 90°$) angle and it can be seen that $\|\vec{x}+\vec{y}\| > \|\vec{x}-\vec{y}\|$. In Figure 1.13, the vectors $\vec{x}$ and $\vec{y}$ form an obtuse ($> 90°$) angle and it can be seen that $\|\vec{x}+\vec{y}\| < \|\vec{x}-\vec{y}\|$. Thus, $\vec{x}$ and $\vec{y}$ are perpendicular to each other if and only if $\|\vec{x}+\vec{y}\| = \|\vec{x}-\vec{y}\|$.

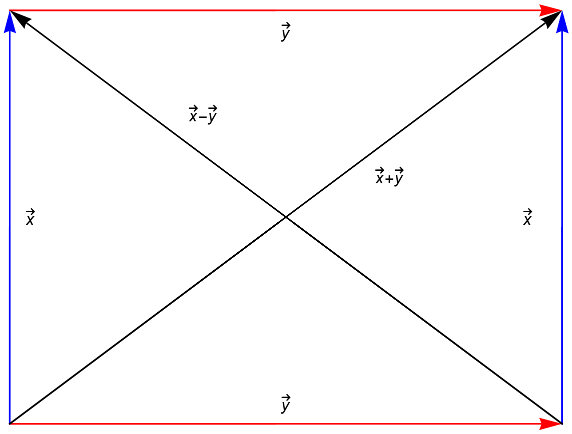


Figure 1.11: $\|\vec{x}+\vec{y}\| = \|\vec{x}-\vec{y}\|$

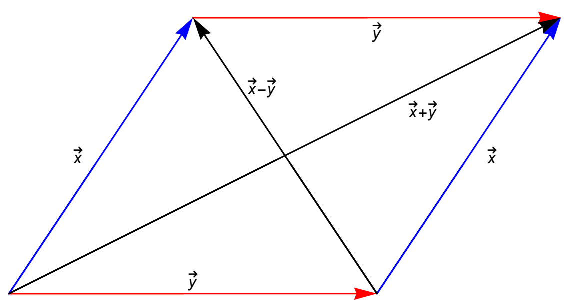


Figure 1.12: $\|\vec{x}+\vec{y}\| > \|\vec{x}-\vec{y}\|$

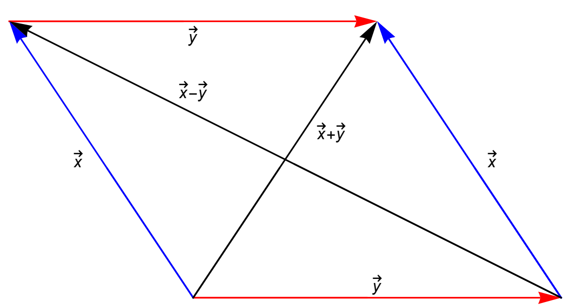


Figure 1.13: $\|\vec{x}+\vec{y}\| < \|\vec{x}-\vec{y}\|$

Having established that $\vec{x}$ and $\vec{y}$ are perpendicular to each other if and only if $\|\vec{x}+\vec{y}\| = \|\vec{x}-\vec{y}\|$, we are prepared to derive the promised algebraic criterion for perpendicularity. We will do this by computing $\|\vec{x}+\vec{y}\|^2$ and $\|\vec{x}-\vec{y}\|^2$ and setting these equal to each other. (When working with vector lengths, it is often convenient to work with squares of lengths rather than with the lengths themselves so that we do not need to deal with square roots throughout our computations.)

Suppose that $\vec{x} = \langle x_1, x_2\rangle$ and $\vec{y} = \langle y_1, y_2\rangle$. Then $\vec{x}+\vec{y} = \langle x_1+y_1, x_2+y_2\rangle$ and we have

$$\begin{aligned}\|\vec{x}+\vec{y}\|^2 &= (x_1+y_1)^2 + (x_2+y_2)^2 \\ &= x_1^2 + 2x_1y_1 + y_1^2 + x_2^2 + 2x_2y_2 + y_2^2 \\ &= \left(x_1^2 + x_2^2\right) + \left(y_1^2 + y_2^2\right) + 2\left(x_1y_1 + x_2y_2\right) \\ &= \|\vec{x}\|^2 + \|\vec{y}\|^2 + 2\left(x_1y_1 + x_2y_2\right).\end{aligned}$$

By a similar calculation we also have

$$\|\vec{x}-\vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2 - 2\left(x_1y_1 + x_2y_2\right).$$

If $\vec{x}$ and $\vec{y}$ are perpendicular to each other, then it must be the case that $\|\vec{x}+\vec{y}\|^2 = \|\vec{x}-\vec{y}\|^2$, and this implies (from the calculations above) that

$$\|\vec{x}\|^2 + \|\vec{y}\|^2 + 2\left(x_1y_1 + x_2y_2\right) = \|\vec{x}\|^2 + \|\vec{y}\|^2 - 2\left(x_1y_1 + x_2y_2\right).$$

The above equation simplifies to

$$4\left(x_1y_1 + x_2y_2\right) = 0,$$

which simplifies to

$$x_1y_1 + x_2y_2 = 0.$$

The quantity $x_1y_1 + x_2y_2$ is called the **dot product** of $\vec{x}$ and $\vec{y}$. It is denoted by $\vec{x}\cdot\vec{y}$. We have discovered that $\vec{x}$ and $\vec{y}$ are perpendicular to each other if and only if $\vec{x}\cdot\vec{y} = 0$.

To illustrate with examples, consider the vectors $\vec{x} = \langle -2, 4\rangle$ and $\vec{y} = \langle 2, 0\rangle$ that are pictured in Figure 1.9. The dot product of these vectors, which are not perpendicular to each other, is

$$\vec{x}\cdot\vec{y} = (-2)(2) + (4)(0) = -4 \neq 0.$$

The vectors $\vec{z} = \langle -2, 2\rangle$ and $\vec{w} = \langle -1, -1\rangle$ pictured in Figure 1.10, which are perpendicular to each other, have dot product

$$\vec{z} \cdot \vec{w} = (-2)(-1) + (2)(-1) = 0.$$

We have defined the concept of perpendicularity by saying that two vectors are perpendicular if their standard representatives form a 90° angle. If one (or both) of the vectors is the zero vector, then this concept does not apply, because there is no well–defined angle between two vectors when one of the vectors is the zero vector. (To determine an angle between two vectors, both of the vectors must have a positive length.) However, for any vector $\vec{x} = \langle x_1, x_2\rangle$ in R^2, we have

$$\vec{x} \cdot \vec{0}_2 = \langle x_1, x_2\rangle \cdot \langle 0, 0\rangle = (x_1)(0) + (x_2)(0) = 0.$$

For this reason, we will define the concept of orthogonality by saying that two vectors, $\vec{x}$ and $\vec{y}$, in R^2 are **orthogonal** to each other if $\vec{x} \cdot \vec{y} = 0$. Thus, two non–zero vectors in R^2 are orthogonal to each other if and only if they are perpendicular to each other and, by definition, the zero vector, $\vec{0}_2$, is orthogonal to every vector in R^2.

Exercise 1.1.15. *Draw pictures of each of the following pairs of vectors, $\vec{x}$ and $\vec{y}$. For each pair, compute $\vec{x} \cdot \vec{y}$ and state whether or not $\vec{x}$ and $\vec{y}$ are orthogonal to each other.*

1. $\vec{x} = \langle 0, -4\rangle$, $\vec{y} = \langle -4, -1\rangle$
2. $\vec{x} = \langle 1, 0\rangle$, $\vec{y} = \langle 3, 3\rangle$
3. $\vec{x} = \langle -4, 6\rangle$, $\vec{y} = \left\langle -1, -\frac{2}{3}\right\rangle$
4. $\vec{x} = \langle -4, 6\rangle$, $\vec{y} = \langle -5, -2\rangle$
5. $\vec{x} = \langle -2, 1\rangle$, $\vec{y} = \langle -3, -6\rangle$
6. $\vec{x} = \langle 1, 5\rangle$, $\vec{y} = \langle 0, 0\rangle$

Exercise 1.1.16. *For any vector $\vec{x} = \langle x_1, x_2\rangle$ in R^2:*

1. *Explain why it must be true that $\vec{x} \cdot \vec{x} \geq 0$.*
2. *Explain why the only possible way to have $\vec{x} \cdot \vec{x} = 0$ is if $\vec{x} = \vec{0}_2$.*

3. *Show that* $\vec{x} \cdot \vec{x} = \|\vec{x}\|^2$.

The dot product is sometimes called a **scalar product** because it is an operation that is computed on two vectors that produces a scalar output. The dot product satisfies select algebraic properties. In particular, the dot product is commutative,

$$\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x},$$

distributes over vector addition,

$$\vec{x} \cdot (\vec{y} + \vec{z}) = \vec{x} \cdot \vec{y} + \vec{x} \cdot \vec{z},$$

and scalar multiplication can be factored out of the dot product,

$$c(\vec{x} \cdot \vec{y}) = (c\vec{x}) \cdot \vec{y} = \vec{x} \cdot (c\vec{y}).$$

These properties can easily be established from the definition of the dot product (see Exercise 1.1.17).

Exercise 1.1.17. *Let* $\vec{x} = \langle x_1, x_2 \rangle$, $\vec{y} = \langle y_1, y_2 \rangle$, *and* $\vec{z} = \langle z_1, z_2 \rangle$ *be any vectors in* R^2 *and* c *be any scalar. Show that*

1. $\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$
2. $\vec{x} \cdot (\vec{y} + \vec{z}) = \vec{x} \cdot \vec{y} + \vec{x} \cdot \vec{z}$
3. $c(\vec{x} \cdot \vec{y}) = (c\vec{x}) \cdot \vec{y} = \vec{x} \cdot (c\vec{y})$

1.1.7 Direction

We have defined a vector in R^2 to be an object whose essence is that it describes a magnitude and a direction. We have formally defined what we mean by the magnitude of a vector. Having given a precise meaning to the concept of "magnitude", we should also give a precise meaning to the concept of "direction". In R^2, we can describe the direction of a given vector by measuring the angles that the standard representative of the vector makes with the coordinate axes. The definition of direction that we give below will be seen to involve these angles. The fact that these angles are involved may not be evident to you as you read the upcoming definition, but we will bring out that fact in the example and exercises that follow.

Definition 1.1.2. *The* ***direction vector*** *of any non–zero vector $\vec{x}$ in R^2 is defined to be the unit vector*

$$\vec{x}_U = \frac{1}{\|\vec{x}\|}\vec{x}.$$

Recall that a unit vector is a vector whose magnitude is 1. To be sure that we understand Definition 1.1.2, note that if $\vec{x}$ is any non–zero vector in R^2, then $\frac{1}{\|\vec{x}\|}$ is a positive scalar, which means that the direction vector $\vec{x}_U = \frac{1}{\|\vec{x}\|}\vec{x}$ points in the same direction as $\vec{x}$. Furthermore,

$$\|\vec{x}_U\| = \left\|\frac{1}{\|\vec{x}\|}\vec{x}\right\| = \frac{1}{\|\vec{x}\|}\,\|\vec{x}\| = 1,$$

and hence $\vec{x}_U$ is a unit vector. Most textbooks do not have a special notation to denote the direction vector that we have denoted by $\vec{x}_U$. We have chosen to use this notation because the subscript "U" reminds us that we are referring to a unit vector (and the $\vec{x}$ part of the notation reminds us that this unit vector points in the same direction as $\vec{x}$).

At first glance, Definition 1.1.2 may seem to be a bit unsatisfactory. We have defined the direction vector of a given vector, $\vec{x}$, to be another vector – in particular the unit vector that points in the same direction as $\vec{x}$. So it almost seems as though the definition we have given is circular – using the concept of direction to define the concept of direction. Our instinct is that we should be defining the concept of direction by measuring certain angles. Indeed, in R^2 (or R^3), we easily could define direction in terms of angles. However, Definition 1.1.2 is the one that most easily generalizes to R^n when $n > 3$, where we find it difficult to envision angles. Let us put our minds at ease about this by looking at an example that illustrates that our definition of direction actually does encapsulate information about angles and agrees with our understanding of trigonometry in R^2.

Example 1.1.2. *The standard representative of the vector $\vec{x} = \langle 3, 7\rangle$ is shown in Figure 1.14. Note that the magnitude of $\vec{x}$ is*

$$\|\vec{x}\| = \sqrt{3^2 + 7^2} = \sqrt{58}.$$

We have labelled the angle θ_1 that the vector $\vec{x}$ makes with the positive x_1 axis (with θ_1 measured from the positive x_1 axis to the vector $\vec{x}$) and we have also *labelled the angle θ_2 that the vector $\vec{x}$ makes with the positive x_2 axis (with θ_2 measured from the positive x_2 axis to the vector $\vec{x}$).*

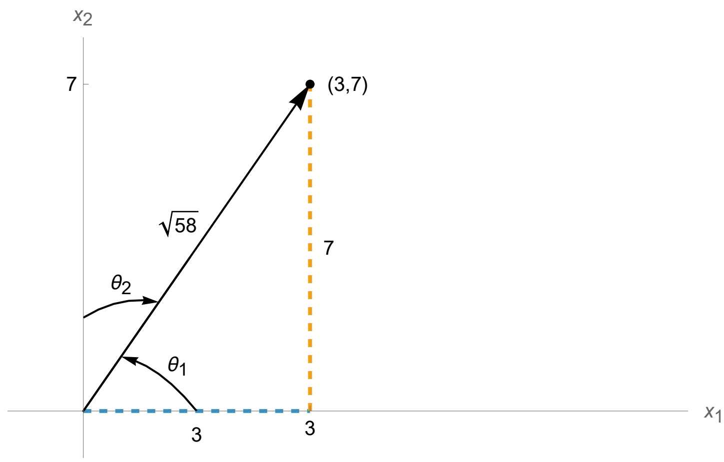


Figure 1.14: The Vector $\vec{x} = \langle 3, 7 \rangle$

By looking at Figure 1.14 and using our knowledge of right triangle trigonometry, we can see that

$$\cos(\theta_1) = \frac{3}{\sqrt{58}}$$

and thus

$$\theta_1 = \cos^{-1}\left(\frac{3}{\sqrt{58}}\right) \approx 66.8^\circ.$$

Likewise,

$$\cos(\theta_2) = \frac{7}{\sqrt{58}}$$

and thus

$$\theta_2 = \cos^{-1}\left(\frac{7}{\sqrt{58}}\right) \approx 23.2^\circ.$$

See Figure 1.15.

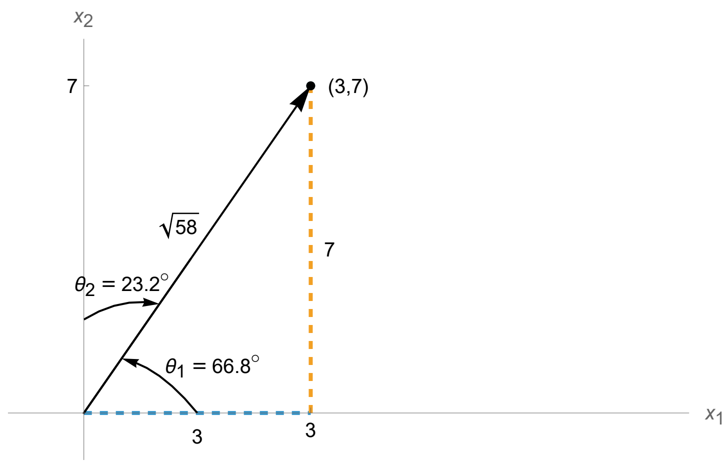


Figure 1.15: Direction Angles of of $\vec{x} = \langle 3, 7\rangle$

The important observation that we wish to make is that

$$\begin{aligned}\cos(\theta_1) &= \frac{3}{\sqrt{58}} = \frac{x_1}{\|\vec{x}\|}\\ \cos(\theta_2) &= \frac{7}{\sqrt{58}} = \frac{x_2}{\|\vec{x}\|}.\end{aligned}$$

Definition 1.1.2 defines the direction vector of the vector $\vec{x} = \langle 3, 7\rangle$ *to be the unit vector*

$$\vec{x}_U = \frac{1}{\|\vec{x}\|}\vec{x} = \frac{1}{\sqrt{58}}\langle 3, 7\rangle = \left\langle \frac{3}{\sqrt{58}}, \frac{7}{\sqrt{58}}\right\rangle$$

and we thus see that the direction vector of $\vec{x}$ *is*

$$\vec{x}_U = \langle \cos(\theta_1), \cos(\theta_2)\rangle .$$

The numbers $\cos(\theta_1)$ *and* $\cos(\theta_2)$ *are called the direction cosines of the vector* $\vec{x}$*. The angles* θ_1 *and* θ_2 *are called the direction angles of* $\vec{x}$*.*

Although we have considered a specific example here, we can be more general. If $\vec{x} = \langle x_1, x_2\rangle$ *is any non–zero vector in* R^2*, then the direction vector of* $\vec{x}$ *is* $\vec{x}_U$ *where*

$$\vec{x}_U = \frac{1}{\|\vec{x}\|}\vec{x} = \frac{1}{\|\vec{x}\|}\langle x_1, x_2\rangle = \left\langle \frac{x_1}{\|\vec{x}\|}, \frac{x_2}{\|\vec{x}\|}\right\rangle = \langle \cos(\theta_1), \cos(\theta_2)\rangle .$$

In the above equation, θ_1 and θ_2 are the direction angles of $\vec{x}$. These are the angles that the standard representative of $\vec{x}$ makes with the positive x_1 and x_2 axes, respectively, assuming that the angles are measured such that the initial side of the angle is at the axis and the terminal side of the angle is $\vec{x}$.

Note that since $\vec{x}_U = \frac{1}{\|\vec{x}\|}\vec{x}$, then

$$\vec{x} = \|\vec{x}\|\,\vec{x}_U.$$

This is a rather nice expression of $\vec{x}$. We originally defined the concept of vector by saying that a vector is an object that describes both a magnitude and a direction. Now that we have given a precise definition of magnitude ($\|\vec{x}\|$) and a precise definition of direction vector ($\vec{x}_U$), we can read the equation $\vec{x} = \|\vec{x}\|\,\vec{x}_U$ as

$$\vec{x} = \textit{(Magnitude of } \vec{x}\textit{) times (Direction vector of } \vec{x}\textit{)}.$$

Motivated by the above example, we will now provide formal definitions of the concepts of direction cosine and direction angle for a vector in R^2.

Definition 1.1.3. *For a non–zero vector $\vec{x} = \langle x_1, x_2\rangle$ in R^2, we define the **direction cosines** of $\vec{x}$ to be the numbers*

$$\frac{x_1}{\|\vec{x}\|} \quad \textit{and} \quad \frac{x_2}{\|\vec{x}\|}.$$

Thus the direction cosines of $\vec{x}$ are the components of the direction vector $\vec{x}_U$.

*We define the **direction angles** of $\vec{x}$ to be the angles*

$$\theta_1 = \cos^{-1}\left(\frac{x_1}{\|\vec{x}\|}\right) \quad \textit{and} \quad \theta_2 = \cos^{-1}\left(\frac{x_2}{\|\vec{x}\|}\right).$$

Remark 1.1.1. *Because the inverse cosine function has range $[0°, 180°]$, the direction angles of a vector in R^2 are always such that $0° \le \theta_1 \le 180°$ and $0° \le \theta_2 \le 180°$.*

Remark 1.1.2. *If we are given the non–zero vector $\vec{x} = \langle x_1, x_2\rangle$, we can compute the magnitude of $\vec{x}$ and the direction cosines of $\vec{x}$. They are*

$$\cos(\theta_1) = \frac{x_1}{\|\vec{x}\|} \quad \textit{and} \quad \cos(\theta_2) = \frac{x_2}{\|\vec{x}\|}.$$

Then we can then find the direction angles of $\vec{x}$ using

$$\theta_1 = \cos^{-1}\left(\frac{x_1}{\|\vec{x}\|}\right) \quad \textit{and} \quad \theta_2 = \cos^{-1}\left(\frac{x_2}{\|\vec{x}\|}\right).$$

Conversely, if we are given the magnitude of $\vec{x}$ and the direction angles of $\vec{x}$, then we can compute $\vec{x}$ using

$$\vec{x} = \|\vec{x}\|\,\vec{x}_U = \|\vec{x}\|\,\langle\cos(\theta_1), \cos(\theta_2)\rangle\,.$$

Exercise 1.1.18. *For each of the following vectors $\vec{x}$ in R^2, draw a picture of the standard representative of $\vec{x}$ and then find*

- *the magnitude of $\vec{x}$*
- *the direction cosines of $\vec{x}$*
- *the direction vector $\vec{x}_U$, of $\vec{x}$, and*
- *the direction angles, θ_1 and θ_2, of $\vec{x}$.*

(Express the angles in degrees – not radians.) Label $\|\vec{x}\|$ and the direction angles, θ_1 and θ_2, in your picture.

1. $\vec{x} = \langle 5, 3\rangle$
2. $\vec{x} = \langle 4, 4\rangle$
3. $\vec{x} = \langle -7, 5\rangle$
4. $\vec{x} = \langle -2, -4\rangle$
5. $\vec{x} = \langle 3, 0\rangle$

Exercise 1.1.19. *In each part below, the magnitude of a vector $\vec{x}$ is given and its direction angles are also given. Draw a picture of the vector. (You should be able to do this just using the given information). Then use the fact that $\vec{x} = \|\vec{x}\|\,\vec{x}_U$ to write $\vec{x}$ in the form $\vec{x} = \langle x_1, x_2\rangle$.*

1. $\|\vec{x}\| = 4$ *and direction angles are* $\theta_1 = 45°$ *and* $\theta_2 = 45°$
2. $\|\vec{x}\| = 2$ *and direction angles are* $\theta_1 = 90°$ *and* $\theta_2 = 0°$

3. $\|\vec{x}\| = 1$ *and direction angles are* $\theta_1 = 30°$ *and* $\theta_2 = 120°$

4. $\|\vec{x}\| = 4$ *and direction angles are* $\theta_1 = 135°$ *and* $\theta_2 = 45°$

5. $\|\vec{x}\| = 2$ *and direction angles are* $\theta_1 = 135°$ *and* $\theta_2 = 135°$

Exercise 1.1.20. *Show that if* $\vec{x} = \langle x_1, x_2\rangle$ *is a non–zero vector in* R^2 *with direction angles* θ_1 *and* θ_2*, then*

$$\cos^2(\theta_1) + \cos^2(\theta_2) = 1.$$

Exercise 1.1.21. *Find the vector* $\vec{y}$ *in* R^2 *that has magnitude* 5 *and points in the same direction as the vector* $\vec{x} = \langle 3, 6\rangle$*.*

1.1.8 Distance Between Vectors

The magnitude of a vector in R^2 corresponds to *length* when we envision a vector as a directed line segment. For the standard representation of $\vec{x} = \langle x_1, x_2\rangle$, the magnitude $\|\vec{x}\|$ is the distance between the terminal point at (x_1, x_2) and the origin. So we can think of $\|\vec{x}\|$ as a measure of the distance between the vector $\vec{x}$ and the zero vector $\vec{0}_2$. We can extend this notion to define the distance between any two vectors $\vec{x}$ and $\vec{y}$ in R^2. If we consider the standard representations of two vectors, $\vec{x} = \langle x_1, x_2\rangle$ and $\vec{y} = \langle y_1, y_2\rangle$, then the difference $\vec{y} - \vec{x}$ has a representation with initial point at (y_1, y_2) and terminal point at (x_1, x_2) as shown in Figure 1.16. We can define the distance between the vectors $\vec{x}$ and $\vec{y}$ by the magnitude of this vector. That is, the distance between $\vec{x} = \langle x_1, x_2\rangle$ and $\vec{y} = \langle y_1, y_2\rangle$ is the distance between their terminal points when considering their standard representations as directed line segments.

Definition 1.1.4. *If* $\vec{x}$ *and* $\vec{y}$ *are vectors in* R^2*, we will denote the distance between the vectors* $\operatorname{dist}(\vec{y}, \vec{x})$*. This distance,*

$$\operatorname{dist}(\vec{y}, \vec{x}) = \|\vec{y} - \vec{x}\|.$$

Remark 1.1.3. *In an elementary algebra setting, you likely would have seen a pair of points denoted by* (x_1, y_1) *and* (x_2, y_2)*, and you would have learned that the distance between them was given by the formula*

$$\textit{Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

We are using a different notational convention here by denoting our points (x_1, x_2) *and* (y_1, y_2)*, but hopefully it is clear that our distance between two vectors matches the familiar formula.*

Exercise 1.1.22. *Show that* $\text{dist}(\vec{y}, \vec{x})$ *is equal to* $\text{dist}(\vec{x}, \vec{y})$ *for any pair of vectors* $\vec{x}$ *and* $\vec{y}$*.*

Exercise 1.1.23. *Find the distance between each set of vectors.*

1. $\vec{x} = \langle 1, 1 \rangle$, $\vec{y} = \langle -2, 1 \rangle$
2. $\vec{x} = \langle 2, 3 \rangle$, $\vec{y} = \langle 0, 0 \rangle$
3. $\vec{x} = \left\langle 2, -\frac{1}{2} \right\rangle$, $\vec{y} = \langle 0, 8 \rangle$
4. $\vec{x} = \langle 1, -1 \rangle$, $\vec{y} = \langle -2, 2 \rangle$

1.2 The Vector Space R^3

Now that we are familiar with the algebraic and geometric structure of the vector space R^2, we can comfortably extend the ideas to include a third component—or dimension. Extending on our definition of a vector in R^2, we define a vector, $\vec{x}$, in R^3 as an ordered triple of real numbers $\vec{x} = \langle x_1, x_2, x_3 \rangle$. As before, we will call the real numbers, x_1, x_2, and x_3, the **entries** or **components** of the vector $\vec{x}$. Figure 1.17 depicts the point $(4, 5, 6)$ and the vector $\vec{x} = \langle 4, 5, 6 \rangle$ in R^3.

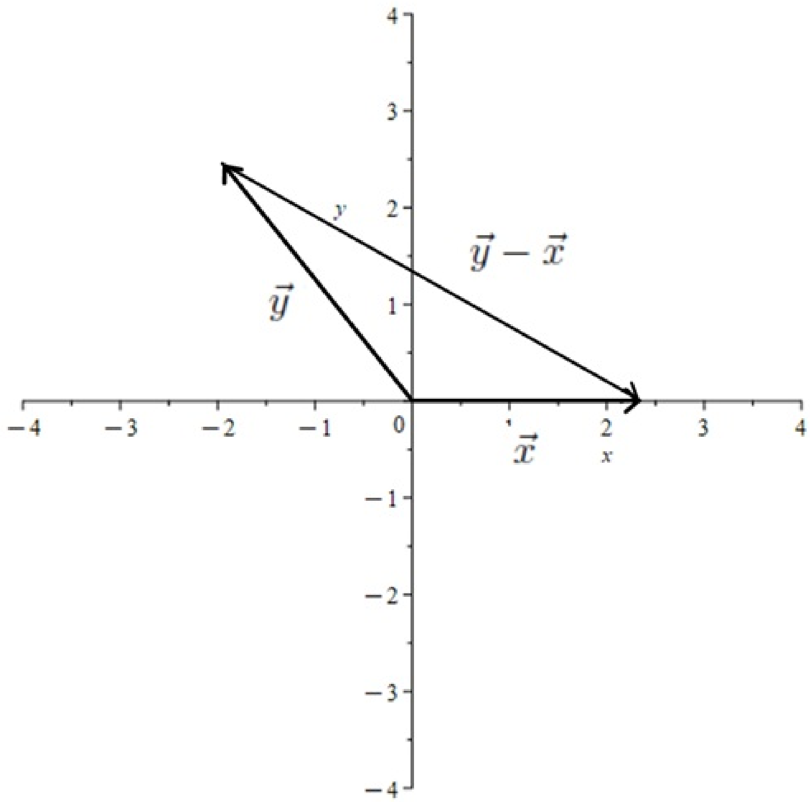


Figure 1.16: The distance between the vectors $\vec{x}$ and $\vec{y}$ is defined by the magnitude of their difference, $\|\vec{y}-\vec{x}\|$ (or $\|\vec{x}-\vec{y}\|$).

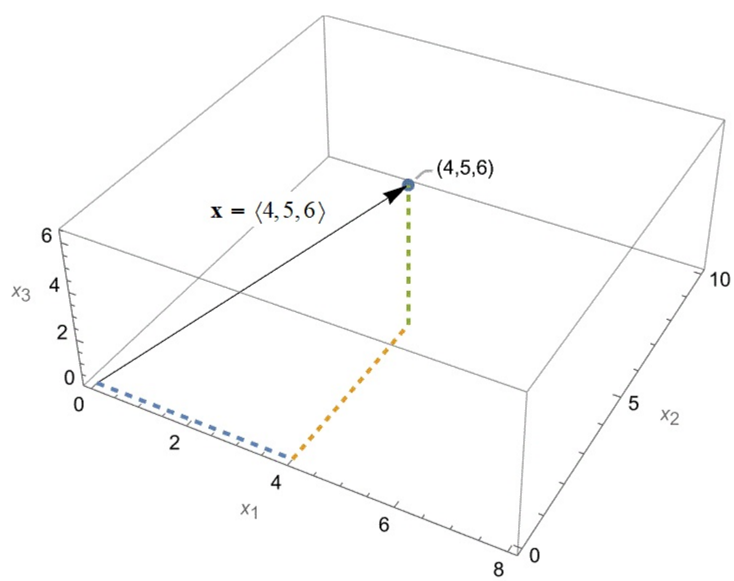


Figure 1.17: The point $(4, 5, 6)$ and the vector $\vec{x} = \langle 4, 5, 6\rangle$ in R^3

Given that we can perceive three spatial dimensions in the world around us (up/down, left/right, forward/back), some geometric intuition can be used when working with vectors in R^3. However, graphs and drawings are two dimensional renderings used to depict three dimensional objects, hence we will primarily rely on algebraic manipulations when we interact with vectors in R^3. We will continue to work with scalars (real numbers) along with vectors in R^3.

As we did in R^2, we will define the two critical operations of **vector addition** and **scalar multiplication**. Letting $\vec{x} = \langle x_1, x_2, x_3 \rangle$ and $\vec{y} = \langle y_1, y_2, y_3 \rangle$ be any vectors in R^3 and c be any scalar, we define

- the vector sum $\vec{x} + \vec{y} = \langle x_1 + y_1, x_2 + y_2, x_3 + y_3 \rangle$, and
- the scalar multiple $c\vec{x} = \langle cx_1, cx_2, cx_3 \rangle$.

The additive identity vector $\vec{0}_3 = \langle 0, 0, 0 \rangle$ is called the **zero vector** in R^3, and given any vector $\vec{x} = \langle x_1, x_2, x_3 \rangle$ in R^3, the vector

$$-\vec{x} = -\langle x_1, x_2, x_3 \rangle = \langle -x_1, -x_2, -x_3 \rangle$$

is its **additive inverse**. As before, we define vector subtraction as vector addition with an additive inverse. Thus

$$\vec{x} - \vec{y} = \vec{x} + (-\vec{y}) = \langle x_1 - y_1, x_2 - y_2, x_3 - y_3 \rangle$$

If we combine vectors in R^3 using the operations of vector addition and scalar multiplication, we refer to the resulting vectors as **linear combinations**.

Example 1.2.1. *Consider the pair of vectors $\vec{x} = \langle 1, 0, 3 \rangle$ and $\vec{y} = \langle -6, 3, 2 \rangle$, in R^3. Evaluate the sum $\vec{x} + \vec{y}$, difference $\vec{x} - \vec{y}$, and the scalar multiplies $3\vec{x}$ and $-2\vec{y}$. Show that $\vec{x} + (-\vec{x}) = \vec{0}_3$.*

We have

- $\vec{x} + \vec{y} = \langle 1 - 6, 0 + 3, 3 + 2 \rangle = \langle -5, 3, 5 \rangle$,
- $\vec{x} - \vec{y} = \langle 1 + 6, 0 - 3, 3 - 2 \rangle = \langle 7, -3, 1 \rangle$,
- $3\vec{x} = 3\,\langle 1, 0, 3 \rangle = \langle 3(1), 3(0), 3(3) \rangle = \langle 3, 0, 9 \rangle$, *and*
- $-2\vec{y} = -2\,\langle -6, 3, 2 \rangle = \langle -2(-6), -2(3), -2(2) \rangle = \langle 12, -6, -4 \rangle$.

Also, note that $-\vec{x} = \langle -1, 0, -3\rangle$ *so that*

$$\vec{x} + (-\vec{x}) = \langle 1-1, 0+0, 3-3\rangle = \langle 0,0,0\rangle = \vec{0}_3.$$

Exercise 1.2.1. *For each pair of vectors* $\vec{x}$ *and* $\vec{y}$*, evaluate* $2\vec{x}$*,* $\vec{x}+\vec{y}$*, and* $\vec{x}-3\vec{y}$*.*

1. $\vec{x} = \langle 1,1,-1\rangle$, $\vec{y} = \langle -2,1,4\rangle$
2. $\vec{x} = \langle 2,3,4\rangle$, $\vec{y} = \langle 0,0,0\rangle$
3. $\vec{x} = \left\langle 4,2,-\frac{1}{2}\right\rangle$, $\vec{y} = \langle 1,0,8\rangle$
4. $\vec{x} = \langle 0,0,-2\rangle$, $\vec{y} = \langle 3,-4,2\rangle$

1.2.1 Magnitude, Dot Product, and Orthogonality

As with R^2, we can identify certain geometric properties associated with vectors in R^3—even if access to illustrations is limited. Similar to our experience in R^2, we can associate the vector $\langle x_1, x_2, x_3\rangle$ with a directed line segment emanating from the origin $(0,0,0)$ in a Cartesian coordinate system (i.e., the xyz-space) and terminating at the point (x_1,x_2,x_3). As such, we can assign the length of such a line segment and consider it the length of the vector $\vec{x} = \langle x_1,x_2,x_3\rangle$. Using the same language and notation we define the **length**, also called the **magnitude**, of a vector

$$\text{Length of}\,\vec{x} = \|\vec{x}\| = \sqrt{x_1^2+x_2^2+x_3^2}.$$

As in R^2, a vector $\vec{x}$ in R^3 having magnitude 1 is called a **unit vector**. The vector $\vec{u} = \left\langle -\frac{1}{3}, \frac{2}{3}, -\frac{2}{3}\right\rangle$ is an example of a unit vector in R^3 because

$$\|\vec{u}\| = \sqrt{\left(-\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2 + \left(-\frac{2}{3}\right)^2} = \sqrt{\frac{1+4+4}{9}} = \sqrt{\frac{9}{9}} = 1.$$

Example 1.2.2. *Find the length of* $\vec{x}$*,* $\vec{y}$ *and* $-\vec{y}$ *where* $\vec{x} = \langle 1,0,3\rangle$ *and* $\vec{y} = \langle -6,3,2\rangle$*. We have*

- $\|\vec{x}\| = \sqrt{1^2+0^2+3^2} = \sqrt{10} \approx 3.162,$
- $\|\vec{y}\| = \sqrt{(-6)^2+3^2+2^2} = \sqrt{49} = 7,$ *and*

- $\| -\vec{y}\| = \sqrt{6^2 + (-3)^2 + (-2)^2} = \sqrt{49} = 7.$

We define the **dot product** of two vectors, $\vec{x} = \langle x_1, x_2, x_3\rangle$ and $\vec{y} = \langle y_1, y_2, y_3\rangle$, in R^3 as

$$\vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + x_3 y_3,$$

and we say that two vectors in R^3 are **orthogonal** if their dot product is zero.

Example 1.2.3. *Determine whether $\vec{x} = \langle 1, 0, 3\rangle$ and $\vec{y} = \langle -6, 3, 2\rangle$ are orthogonal in R^3.*

We compute their dot product

$$\vec{x} \cdot \vec{y} = 1(-6) + 0(3) + 3(2) = 0,$$

and conclude that they are orthogonal.

While the geometry is more complicated in R^3 than in R^2, the dot product of two nonzero vectors has a similar connection to an angle between them (or their standard representations in a coordinate system). Note that for the pair of vectors in Example 1.2.3, we find that

$$\|\vec{x} + \vec{y}\| = \sqrt{(-5)^2 + 3^2 + 5^2} = \sqrt{59} = \sqrt{7^2 + (-3)^2 + 1^2} = \|\vec{x} - \vec{y}\|.$$

This example suggests that a familiar geometric result from R^2 also holds in R^3. Namely, that for any pair of nonzero vectors, $\vec{x}$ and $\vec{y}$, in R^3, $\vec{x} \cdot \vec{y} = 0$, if and only if $\|\vec{x} + \vec{y}\| = \|\vec{x} - \vec{y}\|$. This is in fact true and can be demonstrated algebraically using the same computational approach seen in section 1.1.6. This is left to the reader as Exercise 1.2.3 below.

Exercise 1.2.2. *For each pair of vectors in exercise 1.2.1, determine whether the pair is orthogonal or not orthogonal.*

Exercise 1.2.3. *Let $\vec{x} = \langle x_1, x_2, x_3\rangle$ and $\vec{y} = \langle y_1, y_2, y_3\rangle$. Apply the process used in section 1.1.6 to show that $\|\vec{x} + \vec{y}\|^2 = \|\vec{x} - \vec{y}\|^2$ if and only if $x_1 y_1 + x_2 y_2 + x_3 y_3 = 0$.*

1.2.2 Direction

In Section 1.1.7, we defined the concepts of direction angles, direction cosines, and direction vector of a given non–zero vector in R^2. It was seen that those definitions are sensible because they agree with our understanding of trigonometry. These definitions easily generalize to R^3, where they also agree with trigonometry (although, as has been pointed out, visualization is a bit more difficult in R^3 than in R^2).

Definition 1.2.1. *The* ***direction vector*** *of any non–zero vector $\vec{x}$ in R^3 is defined to be the unit vector*

$$\vec{x}_U = \frac{1}{\|\vec{x}\|}\vec{x}.$$

Definition 1.2.2. *For a non–zero vector $\vec{x} = \langle x_1, x_2, x_3 \rangle$ in R^3, we define the* ***direction cosines*** *of $\vec{x}$ to be the numbers*

$$\frac{x_1}{\|\vec{x}\|}, \quad \frac{x_2}{\|\vec{x}\|}, \quad \text{and} \quad \frac{x_3}{\|\vec{x}\|}$$

Thus the direction cosines of $\vec{x}$ are the components of the direction vector $\vec{x}_U$.

We define the ***direction angles*** *of $\vec{x}$ to be the angles*

$$\theta_1 = \cos^{-1}\left(\frac{x_1}{\|\vec{x}\|}\right), \ \theta_2 = \cos^{-1}\left(\frac{x_2}{\|\vec{x}\|}\right), \quad \text{and } \theta_3 = \cos^{-1}\left(\frac{x_3}{\|\vec{x}\|}\right).$$

Example 1.2.4. *The vector $\vec{x} = \langle 4, 5, 6 \rangle$ is pictured in Figure 1.17. The magnitude of this vector is*

$$\|\vec{x}\| = \sqrt{4^2 + 5^2 + 6^2} = \sqrt{77}.$$

The angle, θ_1, from the positive x_1 axis to the vector $\vec{x}$ satisfies

$$\cos(\theta_1) = \frac{4}{\sqrt{77}}.$$

The angle, θ_2, from the positive x_2 axis to the vector $\vec{x}$ satisfies

$$\cos(\theta_2) = \frac{5}{\sqrt{77}}.$$

The angle, θ_3, from the positive x_3 axis to the vector $\vec{x}$ satisfies

$$\cos(\theta_3) = \frac{6}{\sqrt{77}}.$$

Thus the direction vector of $\vec{x}$ is

$$\vec{x}_U = \frac{1}{\sqrt{77}}\langle 4,5,6\rangle = \left\langle \frac{4}{\sqrt{77}}, \frac{5}{\sqrt{77}}, \frac{6}{\sqrt{77}} \right\rangle$$

The direction angles of $\vec{x}$ are

$$\begin{aligned} \theta_1 &= \cos^{-1}\left(\frac{4}{\sqrt{77}}\right) \approx 62.88^\circ \\ \theta_2 &= \cos^{-1}\left(\frac{5}{\sqrt{77}}\right) \approx 55.26^\circ \\ \theta_3 &= \cos^{-1}\left(\frac{6}{\sqrt{77}}\right) \approx 46.86^\circ. \end{aligned}$$

Exercise 1.2.4. *For each of the following vectors, $\vec{x}$, in R^3, find*

- *the magnitude of $\vec{x}$*
- *the direction cosines of $\vec{x}$*
- *the direction vector, $\vec{x}_U$, of $\vec{x}$, and*
- *the direction angles, θ_1, θ_2, and θ_3, of $\vec{x}$*

1. $\vec{x} = \langle 2,-5,7\rangle$
2. $\vec{x} = \langle 1,0,0\rangle$
3. $\vec{x} = \langle 0,1,0\rangle$
4. $\vec{x} = \langle 1,1,1\rangle$
5. $\vec{x} = \langle 1,2,4\rangle$

Exercise 1.2.5. *Find the vector, $\vec{x} = \langle x_1, x_2, x_3\rangle$ in R^3 that has magnitude $\sqrt{2}$ and direction angles $\theta_1 = 90^\circ$, $\theta_2 = 45^\circ$, and $\theta_3 = 45^\circ$.*

Exercise 1.2.6. *Show that if $\vec{x} = \langle x_1, x_2, x_3\rangle$ is a non–zero vector in R^3 with direction angles θ_1, θ_2, and θ_3, then*

$$\cos^2(\theta_1) + \cos^2(\theta_2) + \cos^2(\theta_3) = 1.$$

Exercise 1.2.7. *Find the vector $\vec{y}$ in R^3 that has magnitude 3 and points in the opposite direction of the vector $\vec{x} = \langle -3,0,4\rangle$.*

1.2.3 Distance Between Vectors in R^3

The distance between vectors in R^3 is defined by the obvious extension of this concept from R^2.

Definition 1.2.3. *For any pair of vectors, $\vec{x}$ and $\vec{y}$, in R^3, the distance between these vectors is denoted* $\text{dist}(\vec{y}, \vec{x})$*, and is defined by*

$$\text{dist}(\vec{y}, \vec{x}) = \|\vec{y} - \vec{x}\|.$$

Exercise 1.2.8. *Find the distance between each pair of vectors.*

1. $\vec{x} = \langle -3, 4, -5 \rangle$, $\vec{y} = \langle 0, 0, 0 \rangle$
2. $\vec{x} = \langle 1, 0, 1 \rangle$, $\vec{y} = \langle 3, -2, 1 \rangle$
3. $\vec{x} = \langle 1, 0, 0 \rangle$, $\vec{y} = \langle 0, 0, 1 \rangle$
4. $\vec{x} = \langle 2, -4, 5 \rangle$, $\vec{y} = \langle 0, 3, 3 \rangle$

Exercise 1.2.9. *Let $\vec{x} = \langle 1, 0, 1 \rangle$ and $\vec{y} = \langle y_1, 3, -2 \rangle$. Find all values of y_1 such that* $\text{dist}(\vec{x}, \vec{y}) = 8$.

1.3 The Vector Spaces R^n In General

We began our discussion on the construction of the vector spaces R^n by focusing exclusively on R^2 – a setting in which our intuition can be guided by drawing pictures to help us understand the main concepts. We then extended these ideas to define vectors in R^3. While drawing pictures is somewhat of an option in R^3 (though it requires more artistic ability—or better yet, a good computer graphics package), we introduced R^3 by extending the algebraic concepts. To extend further to R^n when $n \geq 4$, we mostly have to abandon graphs and pictures altogether, but we can still readily perform the mathematical manipulations that are needed to conceptualize and address problems involving R^n.

Let's recall the various steps we went through in defining and characterizing the vector space R^2. First, we needed two sets of objects and two operations:

- We defined objects called **vectors** having the form $\vec{x} = \langle x_1, x_2 \rangle$ where x_1 and x_2 are real numbers which we called the **entries** (or **components**) of the vector.

- We defined objects called **scalars**. For us, scalars are real numbers (this is the set from which the entries of a vector come).

- We defined the operation **vector addition** allowing us to take two vectors $\vec{x}$ and $\vec{y}$ and form a new vector $\vec{x} + \vec{y}$.

- And we defined **scalar multiplication** allowing us to take a vector $\vec{x}$ and a scalar c and form a new vector $c\vec{x}$.

This gave us the algebraic foundation for R^2. We said that when we use these two operations on vectors, we form **linear combinations**. Then, we added some additional notions and operations related to geometric properties.

- We defined the **magnitude** of a vector in R^2, which we equate with length (since vectors can be associated with line segments), and we defined the **direction** of a vector in R^2 in such a way that the definition is compatible with right triangle trigonometry.

- We defined the **distance** between vectors in R^2 and saw that this corresponded to the distance between points in the plane when those points were the terminal points of standard representations of the vectors, viewed as directed line segments.

- We defined the **dot product** and the property of being **orthogonal**. We saw that with nonzero vectors, the dot product relates to an angle between vectors. In particular, two nonzero, orthogonal vectors are perpendicular.

Our construction of the vector spaces R^n will be analogous to the construction of R^2: Given some integer $n \geq 2$, the vector space R^n will consist of two types of objects on which we define two types of operations.

- We define a set of objects called **vectors**, which have the form

$$\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle$$

  where x_1, x_2, $\ldots$, x_n are real numbers called the **entries** (or **components**) of the vector $\vec{x}$;

- We define objects called **scalars**. As before, scalars are real numbers (this is the set from which the entries of a vector come).

- We define an operation called **vector addition** which is used to add two vectors in R^n to obtain another vector in R^n. For vectors $\vec{x} = \langle x_1, x_2, \ldots, x_n\rangle$ and $\vec{y} = \langle y_1, y_2, \ldots, y_n\rangle$ in R^n, this operation is defined by
$$\vec{x} + \vec{y} = \langle x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n\rangle .$$

- We define an operation called **scalar multiplication** which is used to multiply a vector in R^n by a scalar to obtain another vector in R^n. For a vector $\vec{x} = \langle x_1, x_2, \ldots, x_n\rangle$ in R^n and a scalar c, this operation is defined by
$$c\vec{x} = \langle cx_1, cx_2, \ldots, cx_n\rangle .$$

As we did in R^2, when we use the operations of vector addition and scalar multiplication with vectors in R^n we refer to the result as a linear combination. A formal definition of the concept of a linear combination is given below.

Definition 1.3.1. *Let $S = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ be a set of one or more $(k \geq 1)$ vectors in R^n. A* **linear combination** *of these vectors is any vector of the form*
$$c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_k\vec{x}_k,$$
where $c_1, \ldots, c_k$ are scalars. The coefficients, $c_1, \ldots, c_k$, are often called the **weights**.

Example 1.3.1. *Show that $\vec{v} = \langle 2, 0, 3, -3\rangle$ is a linear combination of $\vec{x}_1 = \langle 1, 0, 0, 0\rangle$ and $\vec{x}_2 = \langle 0, 0, -1, 1\rangle$, and identify the weights.*

We have to show that there are scalars c_1 and c_2 such that $\vec{v} = c_1\vec{x}_1 + c_2\vec{x}_2$. We can set up the equation and then attempt to identify a solution. Note that
$$c_1\vec{x}_1 + c_2\vec{x}_2 = c_1\langle 1, 0, 0, 0\rangle + c_2\langle 0, 0, -1, 1\rangle = \langle c_1, 0, -c_2, c_2\rangle.$$
Comparing this to our vector $\vec{v}$,
$$\langle 2, 0, 3, -3\rangle = \langle c_1, 0, -c_2, c_2\rangle,$$
we see that this requires $c_1 = 2$ and $c_2 = -3$. This demonstrates that $\vec{v}$ is a linear combination of $\vec{x}_1$ and $\vec{x}_2$, and we've found the weights to be $c_1 = 2$ and $c_2 = -3$. In summary,
$$\vec{v} = 2\vec{x}_1 - 3\vec{x}_2.$$

Next, we add notions and operations that provide additional geometric structure to R^n. (Our three-dimensional experience may keep us from drawing pictures of lines and other objects in higher dimensions, but we won't let that stop us for imagining them and discussing things like lengths and angles.)

- We define the **magnitude** of a vector $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle$ in R^n to be
$$\|\vec{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$
- We define the **distance** between two vectors $\vec{x}$ and $\vec{y}$ in R^n to be
$$\text{dist}(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\|.$$
- We define the **dot product** of two vectors $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle$ and $\vec{y} = \langle y_1, y_2, \ldots, y_n \rangle$ in R^n to be
$$\vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$
  Two vectors $\vec{x}$ and $\vec{y}$ in R^n will be said to be **orthogonal** to each other if $\vec{x} \cdot \vec{y} = 0$.

We can equate the magnitude of a vector with length (by imagining a vector in R^n as a directed line segment in some coordinate system with n axes). We will use the term **unit vector** to refer to a vector having magnitude one. As we did in R^2, we define the **direction vector** of a vector $\vec{x} = \langle x_1, x_2. \ldots, x_n \rangle$ in R^n to be the vector

$$\vec{x}_U = \frac{1}{\|\vec{x}\|}\vec{x}$$

and we define the **direction cosines** of $\vec{x}$ to be the numbers

$$\cos(\theta_i) = \frac{x_i}{\|\vec{x}\|},\ i = 1, 2, \ldots, n.$$

The direction cosines of $\vec{x}$ are the components of $\vec{x}_U$. That is

$$\vec{x}_U = \langle \cos(\theta_1), \cos(\theta_2), \ldots, \cos(\theta_n) \rangle.$$

This allows us to write $\vec{x} = \langle x_1, x_2 \ldots, x_n \rangle$ in the convenient form

$$\vec{x} = \|\vec{x}\| \, \vec{x}_U = (\text{Magnitude of } \vec{x}) \text{ times } (\text{Direction vector of } \vec{x}).$$

We can also define the **direction angles** of $\vec{x}$ as

$$\theta_i = \cos^{-1}\left(\frac{x_i}{\|\vec{x}\|}\right),\ i = 1, 2, \ldots, n.$$

However, since most problems that involve the measurement of angles take place in the setting of R^2 or R^3, we usually don't have much need to refer to direction angles in R^n when $n > 3$.

We can say that two nonzero vectors $\vec{x}$ and $\vec{y}$ in R^n are *parallel* to each other if there is a scalar c such that $\vec{y} = c\vec{x}$. And, we can equate orthogonality of nonzero vectors with being *perpendicular* to each other (again, in some plane in some coordinate system with n axes).

Example 1.3.2. *Let $\vec{x} = \langle 1, 2, 1, -1, 0, 4\rangle$ and $\vec{y} = \langle 0, 3, -2, 2, 1, 1\rangle$ be vectors in R^6. Find $\vec{x} + \vec{y}$, $\|\vec{x}\|$, and determine if $\vec{x}$ and $\vec{y}$ are orthogonal.*

Using the operations as defined, we have

$$\vec{x} + \vec{y} = \langle 1 + 0, 2 + 3, 1 - 2, -1 + 2, 0 + 1, 4 + 1\rangle = \langle 1, 5, -1, 1, 1, 5\rangle\,,$$

$$\|\vec{x}\| = \sqrt{1^2 + 2^2 + 1^2 + (-1)^2 + 0^2 + 4^2} = \sqrt{23} \approx 4.7958, \quad \textit{and}$$

$$\vec{x} \cdot \vec{y} = 1(0) + 2(3) + 1(-2) + (-1)(2) + 0(1) + 4(1) = 6.$$

Since $\vec{x} \cdot \vec{y} = 6 \neq 0$, we know that $\vec{x}$ and $\vec{y}$ are not orthogonal vectors in R^6.

Exercise 1.3.1. *For each pair of vectors $\vec{x}$ and $\vec{y}$ in R^n,*

i. identify the value of n,

ii. evaluate $\vec{x} + \vec{y}$ and $\vec{x} - \vec{y}$.

iii. evaluate $\vec{x} \cdot \vec{y}$, and state whether the vectors are orthogonal or not, and

iv. if the pair is orthogonal, confirm that $\|\vec{x} + \vec{y}\| = \|\vec{x} - \vec{y}\|$, and if the pair is not orthogonal, confirm that $\|\vec{x} + \vec{y}\| \neq \|\vec{x} - \vec{y}\|$

v. evaluate $\text{dist}(\vec{x}, \vec{y})$

1. $\vec{x} = \langle 1, -1, 0, 2\rangle$, $\vec{y} = \langle -1, 1, 1, 1\rangle$
2. $\vec{x} = \langle -1, 1, 1, 1, 1, -1\rangle$, $\vec{y} = \langle 0, 1, -1, 2, 0, 0\rangle$
3. $\vec{x} = \langle -3, 0, 4, 1, 2\rangle$, $\vec{y} = \langle 4, 2, 2, 0, 2\rangle$
4. $\vec{x} = \left\langle \frac{1}{3}, -\frac{2}{3}, \frac{2}{3}\right\rangle$, $\vec{y} = \langle 4, 1, -1\rangle$

1.3.1 Algebraic Properties of the Dot Product

The properties of the dot product that we saw in R^2 also hold in R^n for $n \geq 2$. In particular, for any vectors $\vec{x}$, $\vec{y}$ and $\vec{z}$ in R^n and scalar c

- $\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$
- $\vec{x} \cdot (\vec{y} + \vec{z}) = \vec{x} \cdot \vec{y} + \vec{x} \cdot \vec{z}$,
- $\vec{x} \cdot (c\vec{y}) = c(\vec{x} \cdot \vec{y})$,
- $\vec{x} \cdot \vec{x} \geq 0$ with $\vec{x} \cdot \vec{x} = 0$ only if $\vec{x} = \vec{0}_n$.

1.3.2 Span

The operations of vector addition and scalar multiplication will feature prominently through out our study of linear algebra, and when we combine these operations we collectively refer to the result as a linear combination—see Definition 1.3.1. In Section 1.1.5, it was mentioned that we may wish to consider allowing the weights in a linear combination to vary. When we allow the weights (a.k.a. coefficients) in a linear combination to be variable, we obtain what is called a *span*.

Definition 1.3.2. *Let* $S = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ *be a set of one or more* ($k \geq 1$) *vectors in* R^n. *The set of all possible linear combinations of the vectors in* S *is called the* ***span*** *of* S. *It is denoted* $\text{Span}(S)$ *or by* $\text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$.

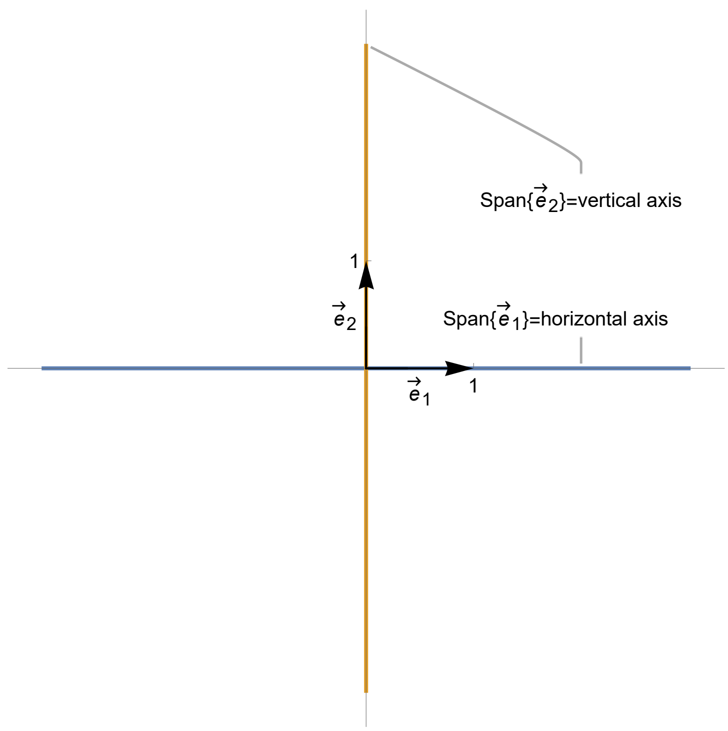
To say that a vector $\vec{y}$ in R^n is in $\text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$, which we can write symbolically as

$$\vec{y} \in \text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\},$$

is to say that there is some set of scalars, $c_1, c_2, \ldots, c_k$, such that

$$\vec{y} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k.$$

To illustrate, let's look back at Example 1.1.1. In that example, we were asked to give a geometric characterization of the collection of all linear combinations of the vector $\vec{e}_1 = \langle 1, 0\rangle$ in R^2. Since a linear combination of this vector is any vector of the form $\langle c, 0\rangle$, we reasoned that we could associate these vectors with all of the points on the horizontal axis in R^2. We can say that $\text{Span}\{\vec{e}_1\}$ is the horizontal axis in R^2. Likewise, for the vector $\vec{e}_2 = \langle 0, 1\rangle$, we can say that $\text{Span}\{\vec{e}_2\}$ is the vertical axis in R^2. See Figure 1.3.2.



Example 1.3.3. *Consider the pair of vectors $\vec{e}_1 = \langle 1, 0\rangle$ and $\vec{e}_2 = \langle 0, 1\rangle$ in R^2. Show that $R^2 = \text{Span}\{\vec{e}_1, \vec{e}_2\}$.*

First, note that any vector in $\text{Span}\{\vec{e}_1, \vec{e}_2\}$, *say*

$$c_1\vec{e}_1 + c_2\vec{e}_2 = \langle c_1, c_2\rangle,$$

is necessarily a vector in R^2. *Next, given any vector* $\vec{x} = \langle x_1, x_2\rangle$ *in* R^2, *we can write*

$$\vec{x} = \langle x_1, 0\rangle + \langle 0, x_2\rangle = x_1\langle 1, 0\rangle + x_2\langle 0, 1\rangle = x_1\vec{e}_1 + x_2\vec{e}_2.$$

It follows that $\vec{x} \in \mathrm{Span}\{\vec{e}_1, \vec{e}_2\}$. *Thus every vector in* R^2 *is in* $\mathrm{Span}\{\vec{e}_1, \vec{e}_2\}$, *and vice versa. It follows that* $R^2 = \mathrm{Span}\{\vec{e}_1, \vec{e}_2\}$.

Remark 1.3.1. *Example 1.3.3 hints at a profound result that we will explore in more depth in Chapter 4. Specifically, it shows that all of* R^2 *can be constructed from a set of building blocks, for example a set of vectors like* $\{\vec{e}_1, \vec{e}_2\}$, *by using the two critical operations. The set* $\{\vec{e}_1, \vec{e}_2\}$ *is particularly easy to work with, and we'll give this a special name in Chapter 3, but it is not the only set of building blocks we can use. For example, we can also say that* $R^2 = \mathrm{Span}\{\vec{u}_1, \vec{u}_2\}$ *where* $\vec{u}_1 = \langle 1, 1\rangle$ *and* $\vec{u}_2 = \langle 1, -1\rangle$. *Note that if* $\vec{x} = \langle x_1, x_2\rangle$ *is any vector in* R^2, *we can write*

$$\vec{x} = \left(\frac{x_1 + x_2}{2}\right)\vec{u}_1 + \left(\frac{x_1 - x_2}{2}\right)\vec{u}_2. \tag{1.1}$$

Exercise 1.3.2. *Verify the claim at the end of Remark 1.3.1. That is, show that the equation (1.1) is true for any vector* $\vec{x} = \langle x_1, x_2\rangle$ *in* R^2, *where* $\vec{u}_1 = \langle 1, 1\rangle$ *and* $\vec{u}_2 = \langle 1, -1\rangle$.

Example 1.3.4. *Consider the three vectors* $\vec{e}_1 = \langle 1, 0, 0\rangle, \vec{e}_2 = \langle 0, 1, 0\rangle$ *and* $\vec{e}_3 = \langle 0, 0, 1\rangle$ *in* R^3. *Describe the subset* $\mathrm{Span}\{\vec{e}_1, \vec{e}_3\}$ *of* R^3.

A vector $\vec{x} \in \mathrm{Span}\{\vec{e}_1, \vec{e}_3\}$ *will have the form*

$$\vec{x} = c_1\vec{e}_1 + c_2\vec{e}_3 = \langle c_1, 0, 0\rangle + \langle 0, 0, c_2\rangle = \langle c_1, 0, c_2\rangle.$$

If we allow the values of c_1 *and* c_2 *to vary over all real numbers, we see that* $\mathrm{Span}\{\vec{e}_1, \vec{e}_3\}$ *contains all vectors in* R^3 *whose second entry is zero. If we equate* R^3 *with Cartesian three-space, i.e., the set of all real triples* (x, y, z)*— then a geometric interpretation of* $\mathrm{Span}\{\vec{e}_1, \vec{e}_3\}$ *is the* xz*-plane.*

Exercise 1.3.3.

1. *If* $\vec{u} = \langle 1, 0, 1\rangle$, *determine whether the following vectors are elements of* $\mathrm{Span}\{\vec{u}\}$.
   (a) $\vec{v} = \langle 2, 0, 2\rangle$
   (b) $\vec{y} = \langle 1, 0, 2\rangle$
   (c) $\vec{0}_3 = \langle 0, 0, 0\rangle$
2. *Suppose* $\vec{x}_1$ *and* $\vec{x}_2$ *are nonzero vectors in* R^n. *Show that the zero vector,* $\vec{0}_n$, *is an element of* $\mathrm{Span}\{\vec{x}_1, \vec{x}_2\}$.

1.4 Additional Exercises

(Jump to Solutions)

1. For $\vec{x}$ in R^n and scalar c in R, use the definition of the magnitude to show that $\|c\vec{x}\| = |c|\|\vec{x}\|$.

2. Consider the vector $\vec{x} = \langle 1, -1, 0, 3\rangle$ in R^4. Determine the value(s) of p such that the vector $\vec{y} = \langle p, 1, 2, p\rangle$ is orthogonal to $\vec{x}$.

3. Let $\vec{x} = \langle -2, 0, 2, 4, 5\rangle$, and $\vec{z} = \langle 4, 6, -3, 2, 2\rangle$. Find a vector $\vec{y}$ in R^5 such that
$$\vec{x} + \vec{y} = \vec{z}$$

4. For each pair of vectors, determine whether they are parallel, orthogonal, or neither parallel nor orthogonal.
   (a) $\vec{x} = \langle 1, -1, 3\rangle$, $\vec{y} = \langle -2, 2, -6\rangle$
   (b) $\vec{x} = \langle 0, 4, 0, -2\rangle$, $\vec{y} = \langle 1, 2, 3, 4\rangle$
   (c) $\vec{x} = \langle 1, 1, 0, 1, 1\rangle$, $\vec{y} = \langle -2, 2, -2, 2, 2\rangle$
   (d) $\vec{x} = \langle 2, -2, 8, 6, 12, 0\rangle$, $\vec{y} = \langle -1, 1, -4, -3, -6, 0\rangle$
   (e) $\vec{x} = \langle 2, 0, -2, 1\rangle$, $\vec{y} = \langle 0, 1, 0, 0\rangle$

5. Let $\vec{x} = \langle 1, 1, 2, 1\rangle$. Find all possible scalars, c such that $\|c\vec{x}\| = 1$.

6. Suppose that the vector $\vec{u}$ in R^n is orthogonal to every other vector in R^n. Explain why it must be that $\vec{u} = \langle 0, 0, \ldots, 0\rangle$. That is, $\vec{u} = \vec{0}_n$, the zero vector in R^n.

7. Let $\vec{u} = \langle -3, 5, 2\rangle$ and $\vec{x} = \langle 1, -1, -4\rangle$. Determine whether $\vec{y} = \langle 0, 1, -5\rangle$ is a linear combination of $\vec{u}$ and $\vec{x}$.

8. Let $\vec{z}_1 = \langle 1, 2\rangle$ and $\vec{z}_2 = \langle 2, 1\rangle$. Show that if $\vec{x} = \langle x_1, x_2\rangle$ is any vector in R^2, then $\vec{x}$ is in Span$\{\vec{z}_1, \vec{z}_2\}$. (Hint: find coefficients c_1 and c_2 such that $\vec{x} = c_1\vec{z}_1 + c_2\vec{z}_2$.)

9. For each statement, indicate whether the statement is true or false. Give a brief explanation or reason for each conclusion.
   (a) If $\vec{x}$ is a vector in R^4 such that $\|\vec{x}\| = 1$, then $\|2\vec{x}\| = 2^4$.

(b) For a vector $\vec{x}$ in R^n, the vector $-\vec{x}$ is the scalar product $-1\vec{x}$.

(c) For any pair of vectors $\vec{x}$ and $\vec{y}$ in R^3, $\|\vec{x}+\vec{y}\| = \|\vec{x}\| + \|\vec{y}\|$.

(d) If a vector $\vec{x}$ in R^n is orthogonal to itself, it must be the zero vector.

(e) If $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is any set of vectors in R^n, then $\vec{0}_n$ is an element of $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$.

(f) If $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is any set of vectors in R^n, then $\vec{0}_n$ is an element of $\text{Span}\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$.

10. Let $\vec{x}$ be any nonzero element of R^5. Explain the difference between the set $\{\vec{x}\}$ and the set $\text{Span}\{\vec{x}\}$.

11. Use the dot product and the fact that $\|\vec{x}\|^2 = \vec{x} \cdot \vec{x}$ to prove the Pythagorean Theorem. The Pythagorean Theorem states

    if $\vec{x}$ and $\vec{y}$ are orthogonal, then $\|\vec{x}+\vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2$.

Chapter 2

Systems of Linear Equations

Central to the study of linear algebra is the desire to solve systems of equations having multiple variables and a special structure known as *linearity*. In fact, the further elements of linear algebra that we consider here (vectors, matrices, vector spaces, linear transformations, etc.) will arise as abstractions motivated by the need to understand and solve systems of linear equations. Let us begin by defining *linear equation* and *system of linear equations*.

2.1 Linear Equations and Linear Systems

The reader may recognize equations of the form $a_1x + a_2y = b$ and $a_1x + a_2y + a_3z = b$ from experience with algebra and geometry and recall that such equations provide an algebraic representation of a line (in R^2) or a plane (in R^3). We understand that the characters x, y, and z represent variables and that a_1, a_2, a_3, and b represent constants. We can call the first a *linear equation in two variables* and the second a *linear equation in three variables*. Here, we won't restrict ourselves to two, three, or any set number of variables, but we can consider these examples prototypes for linear equations. Rather than representing different variables with distinct characters (e.g., x, and y), we will usually use a single character along with subscripts, $x_1, x_2, \dots, x_n$. We'll see that it's no coincidence that this notation matches the notation we used for the entries of a vector in R^n.

Definition 2.1.1. *A* **linear equation** *in the n variables $x_1, x_2, \dots, x_n$ is an equation that can be written in the form*

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

where $a_1, a_2, \ldots, a_n$ and b are real numbers (scalars). The numbers $a_1, \ldots, a_n$ are called the **coefficients***, and b can be called the* **constant term**.

A critical characteristic of a linear equation is that it is constructed by applying exactly two types of operations to the variables: multiplying by scalars, and adding variables. (It's also no coincidence that these two operations, scalar multiplication and vector addition, featured prominently in our discussion of vectors in R^n.) This means that in a linear equation we will not see various other types of operations, such as multiplying variables by one another, raising variables to powers (other than one), and applying trigonometric, exponential or logarithmic functions to them. To illustrate, note that each of

$$4x_1 + 3x_2 - x_3 = 8, \quad \frac{1}{2}x_1 + \sqrt{2}\,x_2 = 0, \quad \text{and} \quad x_1 + x_2 = -2x_3 + 3x_4 + 2$$

is a linear equation, whereas each of

$$\sqrt{2x_1} + x_2 = 3, \quad x_1x_2x_3 = 1, \quad \text{and} \quad \ln(x_1) = x_2 + x_3 + x_4$$

is not a linear equation.

We will define a **system of linear equations** as a collection of one or more linear equations in the same variables considered together.

$$\begin{array}{rcrcrcrcr} 2x_1 & + & x_2 & - & 3x_3 & + & x_4 & = & -3 \\ -x_1 & + & 3x_2 & + & 4x_3 & - & 2x_4 & = & 8 \end{array} \tag{2.1}$$

is an example of a system of two linear equations in four variables.

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & x_3 & = & 0 \\ & + & 3x_2 & - & 2x_3 & = & 3 \\ x_1 & + & x_2 & - & x_3 & = & 3 \end{array} \tag{2.2}$$

is an example of a system of three linear equations in three variables. We can write a generic system consisting of m equations in n variables as

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & + & \vdots & + & \ddots & + & \vdots & = & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}. \tag{2.3}$$

Going forward, we will see that some simple conventions used to write (2.3) will facilitate our work with systems of linear equations. In particular:

- We isolate the constant term in each equation on one side of the equation with all variables on the other.
- We align like variables vertically, even leaving space when a variable does not appear in an equation (e.g., in the second equation of (2.2) where x_1 does not appear).
- When using a double indexed character to represent coefficients, a_{ij}, the first index indicates which equation the coefficient appears in, and the second index corresponds to which variable it scales.

We will classify the system (2.3) as **homogeneous** if each $b_i = 0$. That is, a homogeneous system is one in which every constant term is zero. If at least one $b_i \neq 0$, we will call the system **nonhomogeneous**. While each equation contributes to the system, when analyzing or solving a system, we consider all the equations together as a whole . With this in mind, we define a **solution** and the **solution set** for a system of linear equations.

Definition 2.1.2. *A* **solution** *of (2.3) is as an ordered n-tuple of real numbers, $(s_1, s_2, \ldots, s_n)$, having the property that upon substitution,*

$$x_1 = s_1, \quad x_2 = s_2, \quad \cdots, \quad x_n = s_n,$$

every equation in the system reduces to an identity. The collection of all solutions of (2.3) is called the **solution set** *of the system.*

Definition 2.1.3. *The n-tuple consisting of all zeros, $(0, 0, \ldots, 0)$, is a solution of any homogeneous linear system in n variables. We call this the* **trivial solution**. *A solution, $(s_1, s_2, \ldots, s_n)$ having at least one $s_i \neq 0$ is called a* **nontrivial solution**.

Example 2.1.1. *Consider the linear system (2.2) above. Let us show that $(1, -1, -3)$ is a solution of this system.*

If we set $x_1 = 1$, $x_2 = -1$, and $x_3 = -3$, the equations become

$$\begin{array}{ccccccccccccc} 1 & - & 2(-1) & + & (-3) & = & 1 & + & 2 & - & 3 & = & 0 \\ & + & 3(-1) & - & 2(-3) & = & & - & 3 & + & 6 & = & 3 \\ 1 & + & (-1) & - & (-3) & = & 1 & - & 1 & + & 3 & = & 3 \end{array}.$$

All three equations are satisfied.

In this chapter, we will develop a method for finding all solutions of any system of linear equations - in other words for finding the solution set of the system. It turns out that the solution set of system (2.2) is the set

$$S = \{(2+t, 1+2t, 3t) \mid t \in R\}.$$

Since t *is allowed to be any real number, the system (2.2) has infinitely many solutions. Note that if we set* $t = -1$ *then we obtain the solution* $(1, -1, -3)$. *To verify that every member of the set* S *is a solution of system (2.2), we can set* $x_1 = 2 + t$, $x_2 = 1 + 2t$, *and* $x_3 = 3t$ *while allowing* t *to be an arbitrary real number. Upon substitution, we have*

$$\begin{array}{ccccccccccc} 2+t & - & 2(1+2t) & + & 3t & = & 2-2 & + & (1-4+3)t & = & 0 \\ & + & 3(1+2t) & - & 2(3t) & = & 3 & + & (6-6)t & = & 3 \\ 2+t & + & 1+2t & - & 3t & = & 2+1 & + & (1+2-3)t & = & 3 \end{array}.$$

We see that for any value of the parameter t, *all three equations of system (2.2) are satisfied. Thus every ordered triple in the set* S *is a solution of system (2.2).*

There are three conventions we can use to represent the solution set of a system of linear equations. They are

- set builder notation
- parametric form
- vector parametric form.

We will illustrate each of these methods of representing solution sets using the system of equations (2.2) that was studied in Example 2.1.1.

If we want to describe the solution set of system (2.2) using **set builder notation**, then we say that the solution set is

$$S = \{(2+t, 1+2t, 3t) \mid t \in R\}. \tag{2.4}$$

This is the convention that we used to describe the solution set in Example 2.1.1. The vertical bar (pipe) in the set builder notation is interpreted to mean "such that". The interpretation of (2.4) in words is: "The solution set of the linear system (2.2) is the set of all ordered triples of the form $(2+t, 1+2t, 3t)$ such that t can be any real number."

To represent the solution set of system (2.2) in **parametric form**, we write

$$\begin{aligned} x_1 &= 2 + t \\ x_2 &= 1 + 2t \\ x_3 &= 3t \\ t &\in R. \end{aligned} \tag{2.5}$$

To represent the solution set of system (2.2) in **vector parametric form**, we think of the components (x_1, x_2, and x_3) as being the components of a vector $\vec{x} = \langle x_1, x_2, x_3 \rangle$ and we write

$$\begin{aligned} &\vec{x} = \langle 2 + t, 1 + 2t, 3t \rangle \\ &t \in R. \end{aligned} \tag{2.6}$$

The three ways of describing the solution set are equivalent to each other. The t that appears in all three methods of describing the solution set is called a **parameter**. When using the parametric form or the vector parametric form, we sometimes omit writing "$t \in R$" and take it to be understood that the parameter, t, is allowed to be any real number. However, it is not conventional to omit writing "$t \in R$" when using the set building notation.

Note that since

$$\langle 2 + t, 1 + 2t, 3t \rangle = \langle 2, 1, 0 \rangle + \langle t, 2t, 3t \rangle = \langle 2, 1, 0 \rangle + t\langle 1, 2, 3 \rangle,$$

then an alternative way to write the vector parametric form (2.6) is

$$\vec{x} = \langle 2, 1, 0 \rangle + t\langle 1, 2, 3 \rangle. \tag{2.7}$$

A useful feature of using the form (2.7) is that this form allows us to easily see that the solution set (interpreted as a set of vectors) consists of linear combinations of the vectors $\langle 2, 1, 0 \rangle$ and $\langle 1, 2, 3 \rangle$ where the weight on the vector $\langle 2, 1, 0 \rangle$ is 1 and the vector $\langle 1, 2, 3 \rangle$ can have any (real number) weight.

Exercise 2.1.1.

1. *The solution set of the system*

$$\begin{array}{ccccccccc} 2x_1 & + & 4x_2 & + & 2x_3 & + & 2x_4 & = & -4 \\ x_1 & + & 2x_2 & + & 2x_3 & + & 6x_4 & = & -5 \end{array}$$

has parametric description

$$\begin{array}{rcl} x_1 & = & 1-2s+4t \\ x_2 & = & s, \\ x_3 & = & -3-5t \\ x_4 & = & t, \end{array} \qquad s,t \in R$$

Convert this to vector parametric form.

2. *The solution set of the system*

$$\begin{array}{rcrcrcrcrcr} 3x_1 & + & x_2 & - & 2x_3 & + & 4x_4 & + & 2x_5 & = & -2 \\ x_1 & + & x_2 & + & 2x_3 & - & 2x_4 & + & x_5 & = & -4 \\ 2x_1 & - & x_2 & - & 8x_3 & + & 11x_4 & + & 2x_5 & = & -2 \end{array}$$

is the set of all five-tuples $(x_1, x_2, x_3, x_4, x_5)$ *such that*

$$x_1 = 4 + 2x_3 - 3x_4, \quad x_2 = -2 - 4x_3 + 5x_4, \quad x_5 = -6$$

and x_3 *and* x_4 *can be any real number. Give a parametric description and a vector parametric description of the solution set.*

The solution set of a system is of far more interest than the number of equations it has or the way that it is written. We will say that two systems of equations are **equivalent** if they have the same solution set. Here, we state without proof an important theorem about solutions to systems of linear equations.

Theorem 2.1.1. ***The Solution Set Trichotomy Theorem***
For a system of linear equations, exactly one of the following holds:

i. *The solution set is empty;*

ii. *There exists a unique solution; or*

iii. *There are infinitely many solutions.*

Although we are not able to provide a proof of Theorem 2.1.1 at this point, the reason that the theorem is true will become evident as we proceed through this chapter and develop the tools that are needed to find solution sets of systems of linear equations. In the course of doing this, we will discover that the only three possibilities regarding solutions sets of linear systems are those listed in Theorem 2.1.1. We will refer to a system of equations whose solution set is empty, case i., as **inconsistent**. A system having at least one solution, cases ii. and iii. will be called **consistent**. To distinguish between the two types of consistent systems, case iii. systems are sometimes called **dependent**.

Remark 2.1.1. *The Solution Set Trichotomy Theorem 2.1.1 gives reference to two BIG questions that arise in any problem solving area of mathematics.*

1. *Does a problem even have a solution?*
2. *If a problem has a solution, is that solution unique?*

We call these questions of existence *and* uniqueness*, and we frequently refer back to them.*

Remark 2.1.2. *Every homogeneous system is consistent since the solution set contains at least the trivial solution. Usually, the interesting question when encountering a homogeneous system is whether it also permits nontrivial solutions.*

While Theorem 2.1.1 holds for linear systems of any size, systems having two equations and two variables provide a familiar and intuitive geometric representation of the three solution categories.

2.1.1 Systems of Two Equations with Two Variables

The reader will recognize that an equation of the form $a_1x_1 + a_2x_2 = b$, with at least one of a_1 or a_2 nonzero, can be associated with a line in the x_1x_2-plane. Hence, we can equate a system of two linear equations with two variables (with similar minor assumptions about the coefficients),

$$\begin{array}{ccccc} a_{11}x_1 & + & a_{12}x_2 & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & = & b_2 \end{array},$$

with a pair of lines in R^2. We recall that a pair of lines in the plane will exhibit exactly one of three relationships: they are parallel and never intersect, they will intersect at exactly one point, or they are concurrent (i.e., both equations describe the same line). Since a solution is defined as an ordered pair of real

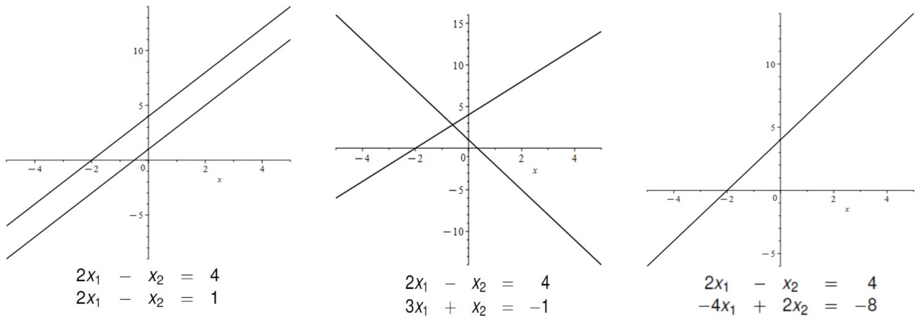


Figure 2.1: Lines determined by two linear equations in two variables illustrating the three possible geometric relationships.

numbers that satisfy both equations in the system, we see that the solution set of a system of two equations in two variables will either be empty (the equations describe parallel lines), consist of a unique ordered pair (the lines have one point of intersection), or will contain infinitely many pairs (all points on the common line). Figure 2.1 illustrates the three types of solution sets stated in Theorem 2.1.1 for systems involving two equations in two variables.

Exercise 2.1.2. *For each system, plot the lines determined by the equations together on the same set of axes and determine whether the system is inconsistent or consistent. If the system is consistent, state whether there is a unique solution or infinitely many solutions.*

1. $$\begin{array}{rcrcr} 3x_1 & + & x_2 & = & 0 \\ x_1 & - & 3x_2 & = & -1 \end{array}$$

2. $$\begin{array}{rcrcr} x_1 & + & x_2 & = & \frac{1}{2} \\ 4x_1 & + & 3x_2 & = & -1 \end{array}$$

3. $$\begin{array}{rcrcl} 4x_1 & + & 6x_2 & = & 3 \\ 6x_1 & + & 9x_2 & = & 0 \end{array}$$

4. $$\begin{array}{rcrcl} 6x_1 & + & 9x_2 & = & 0 \\ 4x_1 & + & 6x_2 & = & 0 \end{array}$$

Exercise 2.1.3. *Consider the system of two equations,*

$$\begin{array}{rcrcl} a_{11}x_1 & + & a_{12}x_2 & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & = & b_2 \end{array} . \tag{2.8}$$

Explain why the system is guaranteed to be consistent with a unique solution whenever $a_{11}a_{22} \neq a_{21}a_{12}$*.(Hint: A pair of lines in the plane are guaranteed to intersect exactly once if they have different slopes.)*

2.2 Solving a System of Linear Equations

In this section, we will consider a methodical approach to solving a linear system of equations. Recall that two systems are called *equivalent* if they have the same solution set. With this in mind, consider the pair of systems

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & - & x_3 & = & 2 \\ 3x_1 & + & x_2 & - & x_3 & = & 2 \\ -x_1 & - & 3x_2 & & & = & -7 \end{array} \quad \text{and} \quad \begin{array}{rcrcrcr} x_1 & + & 2x_2 & - & x_3 & = & 2 \\ & & x_2 & + & x_3 & = & 5 \\ & & & & x_3 & = & 3 \end{array}. \qquad (2.9)$$

Although it is not obvious, these are equivalent. Tasked with finding the solution set, the reader will probably agree that the system on the right has a structure that greatly simplifies that process. In fact, without any effort, we see that any solutions will have to include $x_3 = 3$. This can be substituted into the second equation—we'll call this "back substitution"—to obtain

$$x_2 = 5 - x_3 = 5 - 3 = 2,$$

and with these values known, one more back substitution yields

$$x_1 = 2 - 2x_2 + x_3 = 2 - 2(2) + 3 = 1.$$

We find that the system is consistent and has the unique solution $(1, 2, 3)$. Two features of the system that facilitate this substitution process are the triangular, or inverted stair-step, format of the equations and the fact that the left most variable in each equation has a coefficient of 1. In fact, a third system equivalent to this pair is

$$\begin{array}{ccccc} x_1 & & & = & 1 \\ & x_2 & & = & 2 \\ & & x_3 & = & 3 \end{array}.$$

This formulation has the additional property that, in addition to there being nothing below the left-most variable in each equation, there is nothing above the left-most variable in each equation. We can identify the solution set with no added effort. The critical question is: given a system, how can we obtain an equivalent system with such an advantageous structure?

Given any linear system, there are three operations that we can perform that will preserve the solutions set, i.e., result in an equivalent system. We can

1. multiply an equation by any nonzero constant,
2. interchange the position of any two equations, and
3. replace an equation with the sum of itself and a multiple of any other equation.

We can refer to these operations as 1. scaling, 2. swapping, and 3. replacing. Note that when we refer to the sum of two equations, we mean adding like terms, the common variables and constant terms. To illustrate, consider the system on the left in equation (2.9). If we scale the third equation by the factor -1, the resulting system is

$$\begin{array}{rcrcrcl} x_1 & + & 2x_2 & - & x_3 & = & 2 \\ 3x_1 & + & x_2 & - & x_3 & = & 2 \\ x_1 & + & 3x_2 & & & = & 7 \end{array} . \tag{2.10}$$

Now, in our system (2.10), if we swap the second and third equation, the resulting system is

$$\begin{array}{rcrcrcl} x_1 & + & 2x_2 & - & x_3 & = & 2 \\ x_1 & + & 3x_2 & & & = & 7 \\ 3x_1 & + & x_2 & - & x_3 & = & 2 \end{array} . \tag{2.11}$$

From (2.11), if we replace the third equation with the sum of itself and -3 times the first equation[1], we obtain

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & - & x_3 & = & 2 \\ x_1 & + & 3x_2 & & & = & 7 \\ & - & 5x_2 & + & 2x_3 & = & -4 \end{array} .$$

[1]To see the details, we can line up the third equation and -3 times the first equation and combine like terms

$$\begin{array}{r} \text{third} \\ \text{(-3)first} \\ \text{new third} \end{array} \quad \begin{array}{rcrcrcr} 3x_1 & + & x_2 & - & x_3 & = & 2 \\ -3x_1 & - & 6x_2 & + & 3x_3 & = & -6 \\ \hline 0x_1 & - & 5x_2 & + & 2x_3 & = & -4 \end{array} .$$

It is worth noting here that this sequence of operations has eliminated the variable x_1 from the third equation. If our goal is to obtain the stair-step structure, we have made progress.

2.2.1 Gaussian Elimination

Here, we will describe a methodical process for reducing a given system of equations to one having the desirable structure seen on the right in (2.9). The goal is to use our three operations to eliminate variables (induce zero coefficients), so that we can use back substitution to identify the solution set. This process is referred to as *Gaussian elimination* in honor of German mathematician Carl Friedrich Gauss (1777–1855), though reference to the process can be found in the ancient Chinese mathematics text *The Nine Chapters on the Mathematical Art* composed sometime between the 10th and 2nd century BCE. Rather than attempt to describe the process in the abstract, let's work through an example problem. We will begin with a system of equations and perform a sequence of our three operations. At each step in the process, we can label our equations $E_1, E_2, \ldots$ in the order they appear at that step. We can use the following notation to indicate the operation we choose at that step.

- If we swap equations E_i and E_j, we will write $E_i \leftrightarrow E_j$.
- If we scale equation E_i by the nonzero constant k, we will write $kE_i \rightarrow E_i$.
- If we replace equation E_j with the sum of itself and the number k times equation E_i, we will write $kE_i + E_j \rightarrow E_j$.

Let's begin with the following system of three equations in three variables

$$\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & x_3 & = & 8 \\ x_1 & + & x_2 & + & x_3 & = & 6 \\ x_1 & - & 2x_2 & & & = & -4 \end{array} . \tag{2.12}$$

The general idea is to use the left most variable in the top equation to eliminate this variable from all equations below it. Then, we leave the first equation fixed, move down and use the left most variable in the second equation to eliminate this variable from all equations below it. We continue this process until we have obtained the inverted stair-step format, and we follow up with back substitution to identify the solution set.

It would be advantageous to have the coefficient of x_1 in the top equation to be 1. To this end, let us swap the order of the first and second equation[2]. Recall that we can signify this operation by writing $E_1 \leftrightarrow E_2$. The result is

$$E_1 \leftrightarrow E_2 \qquad \begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 6 \\ 2x_1 & + & x_2 & + & x_3 & = & 8 \\ x_1 & - & 2x_2 & & & = & -4 \end{array} . \tag{2.13}$$

Next, we can eliminate the variable x_1 from the second and third equation with an appropriate replacement. To remove x_1 from the second equation, we need to add $-2x_1$, so we will replace the second equation with the sum of itself and -2 times the first equation. After this operation, we have

$$-2E_1 + E_2 \to E_2 \qquad \begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 6 \\ & - & x_2 & - & x_3 & = & -4 \\ x_1 & - & 2x_2 & & & = & -4 \end{array} . \tag{2.14}$$

Similarly, if we replace the third equation with the sum of itself and -1 times the first equation, we have

$$-E_1 + E_3 \to E_3 \qquad \begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 6 \\ & - & x_2 & - & x_3 & = & -4 \\ & - & 3x_2 & - & x_3 & = & -10 \end{array} . \tag{2.15}$$

At this stage, we have achieved the initial goal. The variable x_1 appears in only the top equation. Let's continue by performing the same process on the subsystem we get by keeping the first equation fixed. That is, we will play this same game on the smaller embedded system consisting of the current second and third equations. Let's scale the second equation by -1 to get a coefficient of 1 on x_2.

$$-1E_2 \to E_2 \qquad \begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 6 \\ & & x_2 & + & x_3 & = & 4 \\ & - & 3x_2 & - & x_3 & = & -10 \end{array} . \tag{2.16}$$

[2]We could also choose to swap the first and third equation or even to scale the first equation by $\frac{1}{2}$. While we'll take a methodical approach, we don't claim that a choice made at a given step is the only choice we could make.

Now, we can eliminate x_2 from the third equation by replacing the third equation with the sum of itself and 3 times the second equation.

$$3E_2 + E_3 \to E_3 \qquad \begin{array}{rcrcrcl} x_1 & + & x_2 & + & x_3 & = & 6 \\ & & x_2 & + & x_3 & = & 4 \\ & & & & 2x_3 & = & 2 \end{array} . \tag{2.17}$$

Finally, we scale the third equation and proceed with the back substitution.

$$\frac{1}{2}E_3 \to E_3 \qquad \begin{array}{rcrcrcl} x_1 & + & x_2 & + & x_3 & = & 6 \\ & & x_2 & + & x_3 & = & 4 \\ & & & & x_3 & = & 1 \end{array} . \tag{2.18}$$

Now, we see that the system has a solution and that $x_3 = 1$. We can perform back substitution to obtain

$$x_2 = 4 - x_3 = 4 - 1 = 3, \quad \text{and} \quad x_1 = 6 - x_2 - x_3 = 6 - 3 - 1 = 2.$$

The system has a unique solution which we can express as an ordered triple $(x_1, x_2, x_3) = (2, 3, 1)$; in parametric form $\begin{array}{rcl} x_1 & = & 2 \\ x_2 & = & 3 \\ x_3 & = & 1 \end{array}$; or in vector parametric form $\vec{x} = \langle 2, 3, 1 \rangle$.

Before moving on, let's go though the process again with the system

$$\begin{array}{rcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & = & 1 \\ 2x_1 & + & x_2 & - & x_3 & = & 2 \\ -x_1 & + & 3x_2 & + & 4x_3 & = & 0 \end{array} . \tag{2.19}$$

We have a coefficient of 1 on x_1 in the first equation. That is advantageous. We can use x_1 in that top equation to eliminate this variable in the second and third. As before, we'll accomplish this with the replacement operations $k_2E_1 + E_2 \to E_2$ and $k_3E_1 + E_3 \to E_3$ with the choices of k_2 and k_3 that result in $0x_1$.

$$-2E_1 + E_2 \to E_2 \qquad \begin{array}{rcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & = & 1 \\ & - & 7x_2 & - & 7x_3 & = & 0 \\ -x_1 & + & 3x_2 & + & 4x_3 & = & 0 \end{array} . \tag{2.20}$$

$$E_1 + E_3 \to E_3 \qquad \begin{array}{rcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & = & 1 \\ & - & 7x_2 & - & 7x_3 & = & 0 \\ & & 7x_2 & + & 7x_3 & = & 1 \end{array} . \tag{2.21}$$

We can scale the second equation to obtain a coefficient of 1 on x_2.

$$-\frac{1}{7}E_2 \to E_2 \qquad \begin{array}{rcrcrcr} x_1 & + & 4x_2 & + & 3x_3 & = & 1 \\ & & x_2 & + & x_3 & = & 0 \\ & & 7x_2 & + & 7x_3 & = & 1 \end{array} . \tag{2.22}$$

Then we can use x_2 in the second equation to eliminate x_2 from the third equation.

$$-7E_2 + E_3 \to E_3 \qquad \begin{array}{rcrcrcr} x_1 & + & 4x_2 & + & 3x_3 & = & 1 \\ & & x_2 & + & x_3 & = & 0 \\ & & & & 0 & = & 1 \end{array} . \tag{2.23}$$

Something interesting has happened; our third equation reads as

$$0x_1 + 0x_2 + 0x_3 = 1, \quad \text{i.e.,} \quad 0 = 1,$$

which is false for all possible values of x_1, x_2, and x_3. It was not initially obvious, but the system (2.19) is inconsistent. Our result, an obviously false equation such as "$0 = 1$," is typical of an inconsistent system.

Exercise 2.2.1. *Perform the Gaussian elimination process on each system of equations. At each step, use the operation notation ($E_i \leftrightarrow E_j$, $kE_i \to E_i$, $kE_i + E_j \to E_j$) to clearly indicate the operation you have selected. If the system is consistent, state the solution in either parametric form or in vector parametric form.*

1. $$\begin{array}{rcrcr} 2x_1 & + & 3x_2 & = & 1 \\ -x_1 & + & 6x_2 & = & -2 \end{array}$$

2. $$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 2x_3 & = & 1 \\ 3x_1 & + & x_2 & - & x_3 & = & -2 \\ x_1 & + & x_2 & - & 2x_3 & = & 0 \end{array}$$

3. $$\begin{array}{rcrcrcr} 3x_1 & + & x_2 & - & x_3 & = & -2 \\ x_1 & + & x_2 & - & 2x_3 & = & 0 \\ 2x_1 & & & + & x_3 & = & 1 \end{array}$$

2.3 Matrices

We might notice in the examples above that, aside from the symbol "$-$" on a negative coefficient, the variable names, and arithmetic symbols "$+$" and "$=$" are carried along at each step as part of the formal expression of the equations while the actual operations affect the coefficients. In this section, we introduce a mathematical object called a *matrix* (plural *matrices*) that will serve as a tool in performing our elimination process. Matrices will allow us to perform the elimination process by focusing on the critical features of the system, the coefficients and the constant terms, while ignoring some of the formal notation.

Definition 2.3.1. *A* **matrix** *is a rectangular array of numbers of the form*

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Each number, a_{ij}, is called an **entry** *or an* **element** *of the matrix. If the matrix has m rows and n columns, we say that the* **size** *or* **dimension** *of the matrix is "m by n" and write $m \times n$.*

Remark 2.3.1. *In writing the size of a matrix, the first number always indicates the number of rows and the second the number of columns. This is the convention that everyone has agreed to, and we will follow suit. In keeping with this tradition, when we use double subscript notation like a_{ij} to indicate an entry in a matrix, the first subscript, i, indicates its row and the second, j, its column.*

Remark 2.3.2. *Matrices are typically labeled using capital letters, A, B, etc., and we often state the size when referring to a matrix. For example, if*

$$A = \begin{bmatrix} 1 & -1 & 3 & 2 \\ 4 & 0 & 2 & 6 \\ -3 & 7 & 1 & 1 \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 3 & 2 \\ -2 & 5 \\ 9 & 1 \end{bmatrix},$$

we say A is a 3×4 matrix and B is a 4×2 matrix. We commonly use an uppercase-lowercase convention when referring to the entries in a matrix.

That is, we would use the notation a_{ij} to refer to the entries in A and b_{ij} to refer to the entries in B. Recalling that the first index indicates the row and the second the columns, we can identify select entries

$$a_{11} = 1, \quad a_{12} = -1, \quad a_{21} = 4, \quad b_{12} = 0, \quad b_{21} = 3, \quad \textit{and} \quad b_{32} = 5.$$

Remark 2.3.3. *The notation $A = [a_{ij}]$ is a common shorthand to refer to a matrix A having entries a_{ij}, especially if the size of the matrix is known or not of immediate interest.*

2.3.1 Coefficient and Augmented Matrices

Given a system of linear equations, we can immediately associate with it a pair of matrices. Recall our generic system (2.3) consisting of m equations in n variables

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

The **coefficient matrix** for the system (2.3) is the $m \times n$ matrix

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

whose entries are the coefficients of the variables in the equations. The number of rows of the coefficient matrix is determined by the number of equations in the system. The number of columns of the coefficient matrix is equal to the number of variables. When we follow the earlier convention for writing a system of linear equations, we can easily identify the coefficient matrix.

The **augmented matrix** for the system (2.3) is the $m \times (n + 1)$ matrix obtained from the coefficient matrix by adding (i.e., *augmenting* it with) an extra column whose entries are the constant terms (right hand side values) of the equations. To distinguish a matrix as the augmented matrix of some system, it is convenient to include a delimiter, such as a dashed or solid line, just to the left of the right most column. While not strictly necessary, it has the advantage of immediately signaling to the reader that the matrix is intended to represent an augmented matrix. If we follow this convention, the augmented matrix for the system (2.3) can be written as

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]. \tag{2.24}$$

Example 2.3.1. *Write the coefficient and augmented matrices for the two systems of equations in (2.9). Those equations were*

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & - & x_3 & = & 2 \\ 3x_1 & + & x_2 & - & x_3 & = & 2 \\ -x_1 & - & 3x_2 & & & = & -7 \end{array} \quad \text{and} \quad \begin{array}{rcrcrcr} x_1 & + & 2x_2 & - & x_3 & = & 2 \\ & & x_2 & + & x_3 & = & 5 \\ & & & & x_3 & = & 3 \end{array}.$$

For the system on the left, the matrices are

$$\text{Coefficient: } \begin{bmatrix} 1 & 2 & -1 \\ 3 & 1 & -1 \\ -1 & -3 & 0 \end{bmatrix}, \quad \text{Augmented: } \left[\begin{array}{rrr|r} 1 & 2 & -1 & 2 \\ 3 & 1 & -1 & 2 \\ -1 & -3 & 0 & -7 \end{array}\right].$$

For the system on the right, the matrices are

$$\text{Coefficient: } \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{Augmented: } \left[\begin{array}{rrr|r} 1 & 2 & -1 & 2 \\ 0 & 1 & 1 & 5 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

Remark 2.3.4. *Note that we place a zero in the matrix corresponding to any position in which a variable is missing. We leave blank spaces when writing a system of equations, but we do not leave blank spaces in a matrix.*

Exercise 2.3.1. *Write the coefficient and the augmented matrix for each system of equations.*

1. $$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 2x_3 & = & 1 \\ 3x_1 & + & x_2 & - & x_3 & = & -2 \\ x_1 & + & x_2 & - & 2x_3 & = & 0 \end{array}$$

2. $$\begin{array}{rcrcrcrcr} -2x_1 & + & x_2 & - & x_3 & + & 2x_4 & = & 1 \\ 6x_1 & - & 3x_2 & & & + & 4x_4 & = & 0 \end{array}$$

Exercise 2.3.2. *For each matrix A, write a homogeneous system of equations having A as its coefficient matrix.*

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

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

Exercise 2.3.3. *For each augmented matrix A, write the corresponding system of equations.*

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

2. $A = \left[\begin{array}{cc|c} 1 & 3 & 5 \\ 7 & 9 & 1 \\ 2 & 4 & 6 \end{array}\right]$

2.3.2 Elementary Row Operations

We can implement the Gaussian elimination procedure on a linear system by performing appropriate operations on the rows of its augmented matrix. There are three such operations that we call **elementary row operations**. The elementary row operations, which can be applied to any matrix, are

1. multiply all entries in a row by any nonzero constant,

2. interchange the position of any two rows, and

3. replace a row with the sum of its entries and a multiple of the corresponding entries in any other row.

We will call these operations *scaling*, *swapping*, and *replacing*, respectively. When we perform one of these operations on a matrix, we say that the resulting matrix is **row equivalent** to the initial matrix. In fact, we will define two matrices as being row equivalent if one can be obtained from the other by performing some sequence of elementary row operations. You may recall that the term "equivalent" was used to describe two systems of linear equations having the same solution set. The following theorem tells us that these two concepts, row equivalence of matrices and equivalence of systems of equations, are intimately related.

Theorem 2.3.1. *If the augmented matrices of two systems of linear equations are row equivalent, then the systems are equivalent.*

This result is not surprising in light of the obvious connection between the elementary row operations and the operations used in Gaussian elimination. But this is a critical result that allows us to use matrices when solving linear systems. Let's revisit our work with the system (2.12) and restate the process in terms of row operations. Going forward, we will use the following popular notation to indicate row operations—fortunately, this matches the previous notation simply replacing E (for equation) with R (for row). We will write

- $kR_i \to R_i$ to indicate scaling the i^{th} row by the constant k,
- $R_i \leftrightarrow R_j$ to indicate swapping rows i and j, and
- $kR_i + R_j \to R_j$ to indicate that row j is replaced with the sum of itself and k times row i.

We start with system (2.12) and write out its augmented matrix.

$$\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & x_3 & = & 8 \\ x_1 & + & x_2 & + & x_3 & = & 6 \\ x_1 & - & 2x_2 & & & = & -4 \end{array} \qquad \left[\begin{array}{rrr|r} 2 & 1 & 1 & 8 \\ 1 & 1 & 1 & 6 \\ 1 & -2 & 0 & -4 \end{array}\right]. \tag{2.25}$$

Let's look at the process we went through before and do a side-by-side comparison of the operations on the equations and the corresponding elementary row operations. See equations (2.13)–(2.18) from section 2.2.1.

$$\overbrace{\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 6 \\ 2x_1 & + & x_2 & + & x_3 & = & 8 \\ x_1 & - & 2x_2 & & & = & -4 \end{array}}^{E_1 \leftrightarrow E_2} \qquad \overbrace{\left[\begin{array}{rrr|r} 1 & 1 & 1 & 6 \\ 2 & 1 & 1 & 8 \\ 1 & -2 & 0 & -4 \end{array}\right]}^{R_1 \leftrightarrow R_2} \tag{2.26}$$

$$\overbrace{\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 6 \\ & - & x_2 & - & x_3 & = & -4 \\ x_1 & - & 2x_2 & & & = & -4 \end{array}}^{-2E_1 + E_2 \to E_2} \qquad \overbrace{\left[\begin{array}{rrr|r} 1 & 1 & 1 & 6 \\ 0 & -1 & -1 & -4 \\ 1 & -2 & 0 & -4 \end{array}\right]}^{-2R_1 + R_2 \to R_2}. \tag{2.27}$$

$$\overbrace{\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 6 \\ & - & x_2 & - & x_3 & = & -4 \\ & - & 3x_2 & - & x_3 & = & -10 \end{array}}^{-E_1+E_3\to E_3} \qquad \overbrace{\left[\begin{array}{rrr|r} 1 & 1 & 1 & 6 \\ 0 & -1 & -1 & -4 \\ 0 & -3 & -1 & -10 \end{array}\right]}^{-R_1+R_3\to R_3}. \tag{2.28}$$

$$\overbrace{\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 6 \\ & & x_2 & + & x_3 & = & 4 \\ & - & 3x_2 & - & x_3 & = & -10 \end{array}}^{-1E_2\to E_2} \qquad \overbrace{\left[\begin{array}{rrr|r} 1 & 1 & 1 & 6 \\ 0 & 1 & 1 & 4 \\ 0 & -3 & -1 & -10 \end{array}\right]}^{-1R_2\to R_2}. \tag{2.29}$$

$$\overbrace{\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 6 \\ & & x_2 & + & x_3 & = & 4 \\ & & & & 2x_3 & = & 2 \end{array}}^{3E_2+E_3\to E_3} \qquad \overbrace{\left[\begin{array}{rrr|r} 1 & 1 & 1 & 6 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 2 & 2 \end{array}\right]}^{3R_2+R_3\to R_3}. \tag{2.30}$$

$$\overbrace{\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 6 \\ & & x_2 & + & x_3 & = & 4 \\ & & & & x_3 & = & 1 \end{array}}^{\frac{1}{2}E_3\to E_3} \qquad \overbrace{\left[\begin{array}{rrr|r} 1 & 1 & 1 & 6 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 1 & 1 \end{array}\right]}^{\frac{1}{2}R_3\to R_3}. \tag{2.31}$$

When we solved this system before, we stopped at this step. Since we can readily convert between an augmented matrix and a system of equations, we could take the matrix on the right in (2.31), write the associated system (the one on the left in (2.31)), and finish up using back substitution. This time, let's continue to use row operations to go further.

Notice that the inverted stair-step structure on the system corresponds to a particular pattern in the augmented matrix. The columns corresponding to the coefficients (the first three in this example) have a sort of triangle of zeros in the lower left corner. There's a special name for this structure, and we'll circle back to that shortly. For now, let's attempt to use our elementary row operations to induce even more elimination. Since we've eliminated x_1 and x_2 from the third equation, we can use row replacement to eliminate x_3 from the first two equations—and we can do this without disrupting that triangle of zeros! Let's see this in action, and look at the corresponding system at each step. We'll perform $-R_3 + R_2 \to R_2$ followed by $-R_3 + R_1 \to R_1$.

$$\overbrace{\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \end{array}\right]}^{-R_3+R_2\to R_2} \qquad \begin{array}{ccccccc} x_1 & + & x_2 & + & x_3 & = & 6 \\ & & x_2 & & & = & 3 \\ & & & & x_3 & = & 1 \end{array}. \tag{2.32}$$

$$\overbrace{\left[\begin{array}{ccc|c} 1 & 1 & 0 & 5 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \end{array}\right]}^{-R_3+R_1\to R_1} \qquad \begin{array}{ccccc} x_1 & + & x_2 & & = 5 \\ & & x_2 & & = 3 \\ & & & x_3 & = 1 \end{array}. \tag{2.33}$$

Finally, we can eliminate x_2 from the first equation using $-R_2 + R_1 \to R_1$ to obtain

$$\overbrace{\left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \end{array}\right]}^{-R_2+R_1\to R_1} \qquad \begin{array}{cccc} x_1 & & & = 2 \\ & x_2 & & = 3 \\ & & x_3 & = 1 \end{array}. \tag{2.34}$$

It is now obvious that the system has a solution and we see what that solution is.

2.3.3 Echelon Forms

If we know what we're looking for, all of the matrices in (2.31)–(2.34) have features that can tell us about the underlying system of equations—features that address those existence and uniqueness questions. A matrix with this structure is called an *echelon form*, more precisely, a *row echelon form*. The word "echelon" is coming from the French échelon, meaning a step or level, which in turn stems from the Latin *scala*, meaning ladder—rather appropriate given our comparison to a set of stairs.

We will call the leftmost nonzero entry in the row of a matrix a **leading entry**. With that, we define what it means to say that a matrix is in row echelon or reduced row echelon form.

Definition 2.3.2. *We will say that a matrix is in* **row echelon form** *if it satisfies the properties that*

1. *any row whose entries are all zeros is below all rows that contain a leading entry, and*

2. *the leading entry in every row is to the right of the leading entries in every row above it.*

We will say that a matrix is in **reduced row echelon form** *if, in addition to being in row echelon form,*

3. *the leading entry in each row is a* 1 *(called a "leading one"), and*

4. *each leading one is the only nonzero entry in its column.*

We'll often use the shorthand "ref" and "rref," respectively, to refer to row echelon and reduced row echelon forms[3]. At first pass, the properties in definition 2.3.2 might seem rather unintuitive, but with a little bit of practice, it's easy to train one's eye to recognize echelon forms, and we will use them extensively. Let's practice identifying echelon forms.

Example 2.3.2. *Let $\star$ represent some nonzero number and $\square$ represent any number (including zero). List all of the possible 2×2 echelon forms.*

There are four possible 2×2 echelon forms.

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \star & \square \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & \star \\ 0 & 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \star & \square \\ 0 & \star \end{bmatrix}.$$

Example 2.3.3. *Classify each of the following matrices as a row echelon form (ref), a reduced row echelon form (rref), or not an echelon form.*

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 0 & -1 & 1 \\ 0 & 0 & 7 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

$$D = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} 1 & 2 & 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 & 6 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

A: The matrix A is in row echelon form, but it is not in reduced echelon form. There are no rows of all zero, so that property is satisfied by default. The leading entries are $2, -1$, and 7, in that order. The leading entries tend strictly to the right as we go down the rows.

[3]Many computer algebra systems capable of matrix manipulations use the names ref and rref for predefined functions that input a matrix and output a row equivalent echelon form.

B: *Matrix B is not an echelon form. We must be careful here because at first glance, B has a bit of a stair step structure. But notice that the leading entry in the third row is not to the right of the leading entry in the second row. An alternative formulation of property 2. in definition 2.3.2 is that all the entries below a leading entry must be zero. Since the leading entry in the second row has a nonzero entry below it, property 2. is violated.*

C: *Matrix C is a reduced row echelon form. It is an echelon form (properties 1. and 2.). Moreover, the leading entries are both 1, and each leading 1 is the only nonzero entry in its column.*

D: *Matrix D is a row echelon form, but it is not a reduced row echelon form. It satisfies properties 1., 2., and 3. But note that the leading one in the second column is not the only nonzero entry in that column.*

E: *Matrix E is a reduced row echelon form. This example is a bit less obvious because it has several nonzero, and non-one, entries. However, if we look closely, we see that all of the properties of an rref are satisfied. This matrix has three leading ones columns 1, 3, and 4. Each leading one is to the right of all leading ones above it, and each leading one is the only nonzero entry in its column.*

Exercise 2.3.4. *Classify each matrix as a row echelon form (ref), a reduced row echelon form (rref), or not an echelon form. Identify which property (or properties) is not satisfied if a matrix is not an echelon form (or is an ref but not an rref).*

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

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

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

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

Exercise 2.3.5. *Use the notation from Example 2.3.2 where appropriate.*

1. *Write out all possible* 2×2 *reduced row echelon forms.*
2. *Write out all of the possible* 2×3 *row echelon forms.*
3. *Write out all of the possible* 2×3 *reduced row echelon forms.*
4. *Write out all possible* 3×3 *row echelon forms.*
5. *Write out all possible* 3×3 *reduced row echelon forms.*

2.3.4 Row Reduction

Given any matrix that is not an echelon form, we can obtain a row equivalent echelon form by performing some set of elementary row operations. We'll call this process **row reduction**. With an echelon form as the goal of our row reduction efforts, we want to choose operations carefully at each step—not because there is only *one correct set of steps*, but because we don't want to choose an operation that takes us further from our goal. We can follow some simple guidelines to optimize our row reduction efforts. In particular, we will work from left to right, top down, to obtain an ref by inducing zeros below each leading entry. To obtain an rref, we first obtain an ref, and then continue the process of inducing zeros above each leading entry by working from right to left, bottom-up, and eventually scaling each leading entry to be 1. Let's see the process in action.

We will start with the 4×5 matrix

$$A = \begin{bmatrix} 3 & 2 & 1 & 6 & 0 \\ 4 & 2 & 2 & 0 & -2 \\ 1 & 1 & 0 & 3 & -2 \\ 2 & 1 & 1 & 3 & 2 \end{bmatrix}$$

and produce an ref. Since the first column is not all zero, the top left corner will contain a leading entry, and we use this leading entry to obtain all zero below it. Letting $\star$ denote a nonzero entry, we want the first column to become $\begin{bmatrix} \star \\ 0 \\ 0 \\ 0 \end{bmatrix}$. It is desirable (though not necessary) to have a 1 in that top left corner, so we will swap the first and third row to obtain

$$R_1 \leftrightarrow R_3 \qquad \begin{bmatrix} 1 & 1 & 0 & 3 & -2 \\ 4 & 2 & 2 & 0 & -2 \\ 3 & 2 & 1 & 6 & 0 \\ 2 & 1 & 1 & 3 & 2 \end{bmatrix}.$$

To clear the column, we use replacement operations $k_1R_1 + R_2 \to R_2$, $k_2R_1 + R_3 \to R_3$ and $k_3R_1 + R_4 \to R_4$ choosing the scalars k_i to that the first entry in the new row is zero. (Hopefully it's clear why moving the 1 to the top row was advantageous.) For this example, we can do three replacement operations.

$$-4R_1 + R_2 \to R_2 \qquad \begin{bmatrix} 1 & 1 & 0 & 3 & -2 \\ 0 & -2 & 2 & -12 & 6 \\ 3 & 2 & 1 & 6 & 0 \\ 2 & 1 & 1 & 3 & 2 \end{bmatrix},$$

$$-3R_1 + R_3 \to R_3 \qquad \begin{bmatrix} 1 & 1 & 0 & 3 & -2 \\ 0 & -2 & 2 & -12 & 6 \\ 0 & -1 & 1 & -3 & 6 \\ 2 & 1 & 1 & 3 & 2 \end{bmatrix}, \quad \text{and}$$

$$-2R_1 + R_4 \to R_4 \qquad \begin{bmatrix} 1 & 1 & 0 & 3 & -2 \\ 0 & -2 & 2 & -12 & 6 \\ 0 & -1 & 1 & -3 & 6 \\ 0 & -1 & 1 & -3 & 6 \end{bmatrix}.$$

Now that the first column has the correct format, we essentially ignore the top row and leftmost column,

$$\begin{bmatrix} 1 & 1 & 0 & 3 & -1 \\ 0 & -2 & 2 & -12 & 6 \\ 0 & -1 & 1 & -3 & 6 \\ 0 & -1 & 1 & -3 & 6 \end{bmatrix},$$

and repeat the process on the resulting sub-matrix (the blue entries above). Leaving the top row fixed, the zeros in the leftmost column guarantee that we will not undo the progress we've made.

Since the leftmost blue column is not all zero, the top entry will be a leading entry and we choose operations that result in the leftmost blue column becoming $\begin{bmatrix} \star \\ 0 \\ 0 \end{bmatrix}$. We can accomplish that via a few steps[4]

$$-\frac{1}{2}R_2 \to R_2 \qquad \begin{bmatrix} 1 & 1 & 0 & 3 & -2 \\ 0 & 1 & -1 & 6 & -3 \\ 0 & -1 & 1 & -3 & 6 \\ 0 & -1 & 1 & -3 & 6 \end{bmatrix},$$

$$R_2 + R_3 \to R_3 \qquad \begin{bmatrix} 1 & 1 & 0 & 3 & -2 \\ 0 & 1 & -1 & 6 & -3 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & -1 & 1 & -3 & 6 \end{bmatrix},$$

$$R_2 + R_4 \to R_4 \qquad \begin{bmatrix} 1 & 1 & 0 & 3 & -2 \\ 0 & 1 & -1 & 6 & -3 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 3 & 3 \end{bmatrix}.$$

Now we move down and to the right, ignoring the first two rows and columns

$$\begin{bmatrix} 1 & 1 & 0 & 3 & -1 \\ 0 & 1 & -1 & 6 & -3 \\ 0 & 0 & \color{blue}{0} & \color{blue}{3} & \color{blue}{3} \\ 0 & 0 & \color{blue}{0} & \color{blue}{3} & \color{blue}{3} \end{bmatrix},$$

and work with the new sub-matrix (shown in blue). We will continue this process until an ref is obtained.

We note that the leftmost column in the new submatrix contains all zeros. There are no row operations that will result in nonzero numbers in these positions without losing the progress we made on the first two columns. This column **will not contain a leading entry**, so we leave those zeros

[4] Keep in mind that it's not the exact operations performed, it's the result that matters. You might choose different operations to achieve the same outcome—and that's perfectly legit!

and move to the right.

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

The goal is for that leftmost blue column to have the form $\begin{bmatrix} \star \\ 0 \end{bmatrix}$. We can accomplish that with one row operation.

$$-R_3 + R_4 \to R_4 \qquad \begin{bmatrix} 1 & 1 & 0 & 3 & -2 \\ 0 & 1 & -1 & 6 & -3 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{2.35}$$

Now we have an ref, and we can see that the first, second and fourth columns in this example contain a leading term. While this ref is not unique (we could, for example, scale any of the rows and still have an ref), further row operations will not change the locations of the leading terms.

We can obtain an rref starting from an ref by essentially working backwards (right to left, bottom up) to obtain zeros in every entry above each leading entry. Performing the process from right to left guarantees that once we get a zero in a desired position, it will remain zero as we proceed.

Continuing with our example, we can obtain zeros above the leading entry in the fourth column by performing two replacements of the form $k_1R_3+R_2 \to R_2$ and $k_2R_3 + R_1 \to R_1$ with strategic choices of k_1 and k_2.

$$-2R_3 + R_2 \to R_2 \qquad \begin{bmatrix} 1 & 1 & 0 & 3 & -2 \\ 0 & 1 & -1 & 0 & -9 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

$$-R_3 + R_1 \to R_1 \qquad \begin{bmatrix} 1 & 1 & 0 & 0 & -5 \\ 0 & 1 & -1 & 0 & -9 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Now we move left to the next leading entry and use replacements to get all zero entries above it. In this example, the next leading entry is in the second column, and we only need one replacement to eliminate the entry above it.

$$-R_2 + R_1 \to R_1 \qquad \begin{bmatrix} 1 & 0 & 1 & 0 & 4 \\ 0 & 1 & -1 & 0 & -9 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

If necessary, we scale all leading entries to be 1, and the process is complete.

$$\frac{1}{3}R_3 \to R_3 \qquad \begin{bmatrix} 1 & 0 & 1 & 0 & 4 \\ 0 & 1 & -1 & 0 & -9 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{2.36}$$

The matrix above in equation (2.36) is an rref that is row equivalent to the matrix A that we started with. We might call this "the reduced row echelon form of A," but such a statement immediately raises a question:

> If, starting from the same matrix A, we select a different set of row operations to obtain an rref, could we obtain a different rref?

This is an interesting question, especially in light of the observation that the ref in equation (2.35) is not unique. But, while we can do row operations to obtain different refs, one key property would remain the same. The positions of the leading entries would not change. Turns out, no matter what operations we choose to obtain a reduced row echelon form, we end up with the same final result. We have the following theorem, stated here without proof.

Theorem 2.3.2. *A matrix A is row equivalent to exactly one reduced row echelon form.*

Since an rref is unique, we can use the notation $\text{rref}(A)$ to refer to the reduced row echelon form of a matrix A.

Exercise 2.3.6. *For each matrix A, follow the process outlined in the row reduction example to find* $\text{rref}(A)$.

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

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

3. $A = \begin{bmatrix} 2 & 4 & 6 & 8 \\ 4 & 6 & 8 & 10 \\ 6 & 8 & 10 & 4 \end{bmatrix}$

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

5. $A = \begin{bmatrix} 1 & 0 & -3 & 0 & 0 \\ 1 & 8 & -5 & -2 & 0 \\ 1 & 6 & -6 & 0 & -1 \\ 3 & 7 & -7 & -1 & -2 \end{bmatrix}$

The leading entries are used to eliminate all of the entries in their respective columns during the row reduction process, and the location of the leading entries is independent of the specific row operations chosen. Given their role, the leading entries are often called **pivots**. For a coefficient or an augmented matrix, we can relate these entries to the variables in the underlying system of equations. Since the reduced row echelon form is unique, we can make the following unambiguous definition.

Definition 2.3.3. *A* **pivot position** *in a matrix A is the location of a leading* 1 *in* $\text{rref}(A)$*. A column that contains a pivot position is called a* **pivot column**.

Example 2.3.4. *Circle the pivot positions and list the pivot columns of the matrix*

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

We previously found

$$\text{rref}(A) = \begin{bmatrix} 1 & 0 & 1 & 0 & 4 \\ 0 & 1 & -1 & 0 & -9 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

From the locations of the leading ones, we identify the entries in the pivot positions. Circling them we have

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

We see that the pivot columns are columns 1, 2, and 4.

Exercise 2.3.7. *For each matrix A in exercise 2.3.6, circle the pivot positions and list the pivot columns.*

Exercise 2.3.8. *Suppose A is a 5×7 matrix.*

1. *If A is the coefficient matrix of a linear system of equations, how many variables does the system have?*

2. *If A is the augmented matrix of a linear system of equations, how many variables does the system have?*

3. *Could A have 7 pivot columns? (Explain your answer.)*

Exercise 2.3.9. *If A is an $m \times n$ matrix, what is the maximum number of pivot columns A can have? (Hint: consider the possible cases, $m < n$ and $m \geq n$. Explain your answer.)*

2.4 Solutions of Linear Systems

Augmented matrices and row reduction provide a convenient framework for solving systems of linear equations, and we'll find that we can use pivot columns to deduce consistency and express solutions. To illustrate, let's consider the system of four equations in five variables

$$\begin{array}{rcrcrcrcrcr} -x_1 & - & 2x_2 & & & & x_4 & + & 2x_5 & = & -3 \\ x_1 & + & 2x_2 & + & x_3 & & & - & 3x_5 & = & 4 \\ 3x_1 & + & 6x_2 & + & x_3 & - & x_4 & - & 4x_5 & = & 10 \\ 2x_1 & + & 4x_2 & + & x_3 & - & x_4 & - & 5x_5 & = & 7 \end{array}. \tag{2.37}$$

The system is formatted nicely, so we can easily write out the augmented matrix,

$$\left[\begin{array}{rrrrr|r} -1 & -2 & 0 & 1 & 2 & -3 \\ 1 & 2 & 1 & 0 & -3 & 4 \\ 3 & 6 & 1 & -1 & -4 & 10 \\ 2 & 4 & 1 & -1 & -5 & 7 \end{array}\right]. \tag{2.38}$$

Now, if we perform the row reduction procedure on this matrix (exercise left to the reader), we find the rref

$$\left[\begin{array}{rrrrr|r} 1 & 2 & 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 0 & -4 & 1 \\ 0 & 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]. \tag{2.39}$$

We can translate this matrix back into a linear system. Since the fourth row is all zero, it corresponds to the trivially true equation "$0 = 0$," which we won't bother to write out. Otherwise, the matrix in (2.39) is the augmented matrix of the system

$$\begin{array}{ccccccccc} x_1 & + & 2x_2 & & & & + & x_5 & = & 3 \\ & & & & x_3 & & - & 4x_5 & = & 1 \\ & & & & & x_4 & + & 3x_5 & = & 0 \end{array}. \tag{2.40}$$

We might notice something interesting about the variables that correspond to the pivot columns. In this case, the pivot columns are columns 1, 3, and 4. Since the only nonzero entry in a pivot column is 1, each of the variables x_1, x_3 and x_4 only appears in one equation, and each has a coefficient of 1. This provides us with a convenient way to express solutions by simply moving the remaining variables (the non-pivot column variables) to the right side. We can write

$$\begin{array}{rcl} x_1 & = & 3 - 2x_2 - x_5 \\ x_3 & = & 1 + 4x_5 \\ x_4 & = & -3x_5 \end{array}. \tag{2.41}$$

There are no additional conditions on the variables x_2 and x_5 which means that these variables can take on any real value. Then, for any choice of x_2 and x_5, as long as x_1, x_3, and x_4 satisfy the three equations (2.41), the 5-tuple $(x_1, x_2, x_3, x_4, x_5)$ will be in the solution set of the system (2.37). Variables like x_2 and x_5 that can take on any value are called **free variables**. It is customary (though not required) to assign parameter names, such as t or s, to free variables. Whether we relabel the free variables or let them keep their original names, a proper presentation of the solution set should clearly identify any free variable(s). A parametric representation of the solution set of (2.37) could be written as

$$\begin{array}{rcl} x_1 & = & 3-2t-s \\ x_2 & = & t \\ x_3 & = & 1+4s \\ x_4 & = & -3s \\ x_5 & = & s \end{array} \quad , \qquad s,t \in R. \tag{2.42}$$

This solution expressed in vector parametric form is

$$\vec{x} = \langle 3,0,1,0,0\rangle + t\langle -2,1,0,0,0\rangle + s\langle -1,0,4,-3,1\rangle. \tag{2.43}$$

As long as we identify free variables, we can also express the solution set in the form

$$\begin{array}{ccl} x_1 & = & 3-2x_2-x_5 \\ x_3 & = & 1+4x_5 \\ x_4 & = & -3x_5 \\ x_2, x_5 & \text{are} & \text{free} \end{array} .$$

Note that the only real difference between this and (2.41) is that we explicitly state that the variables x_2 and x_5 are free. We'll call the variables that are not free **basic variables**. We can formally define the two variable types.

Definition 2.4.1. *Let A be an $m \times n$ matrix that is the coefficient matrix for a system of linear equations in the n variables, $x_1, x_2, \ldots, x_n$. For each $i = 1, \ldots, n$*

- *if the i^{th} column of A is a pivot column, then x_i is a* **basic variable***, and*
- *if the i^{th} column of A is not a pivot column, the x_i is a* **free variable***.*

When expressing the solution of a consistent linear system, we will always express the basic variables in terms of the free variables (and never the other way around[5]).

[5]It's not technically wrong to rearrange the third equation in (2.41) to read $x_5 = -\frac{1}{3}x_4$.

Example 2.4.1. *Use an augmented matrix and row reduction to determine the solution set of the linear system*

$$\begin{array}{ccccccc} 2x_1 & + & 4x_2 & + & 6x_3 & = & 8 \\ 4x_1 & + & 6x_2 & + & 8x_3 & = & 10 \\ 6x_1 & + & 8x_2 & + & 10x_3 & = & 4 \end{array}.$$

We set up the augmented matrix and row reduce to an rref.

$$\left[\begin{array}{ccc|c} 2 & 4 & 6 & 8 \\ 4 & 6 & 8 & 10 \\ 6 & 8 & 10 & 4 \end{array}\right] \xrightarrow{rref} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

The third row of the rref corresponds to the equation

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

which is false for all possible values of x_1, x_2, and x_3. The conclusion is that this system is inconsistent. We can say that the solution set is empty.

Example 2.4.2. *Use an augmented matrix and row reduction to determine the solution set of the linear system*

$$\begin{array}{ccccccc} -x_1 & + & 2x_2 & + & 4x_3 & = & 3 \\ 3x_1 & + & x_2 & + & 2x_3 & = & 4 \\ & - & 2x_2 & + & 6x_3 & = & 1 \end{array}.$$

Again, we set up the augmented matrix and row reduce to an rref.

$$\left[\begin{array}{ccc|c} -1 & 2 & 4 & 3 \\ 3 & 1 & 2 & 4 \\ 0 & -2 & 6 & 1 \end{array}\right] \xrightarrow{rref} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 5/7 \\ 0 & 1 & 0 & 32/35 \\ 0 & 0 & 1 & 33/70 \end{array}\right]$$

The first three columns are pivot columns, so all three variables are basic variables. This system has a unique solution that we can state in parametric form

$$\begin{array}{ccc} x_1 & = & 5/7 \\ x_2 & = & 32/35 \\ x_3 & = & 33/70 \end{array}.$$

But only one of x_4 and x_5 can truly be a free parameter. We would have to rearrange the other two equations in (2.41) to replace each x_5 with $-\frac{1}{3}x_4$. That's extra work without good reason for doing it. It would be incorrect to say both $x_4 = -3x_5$ and $x_5 = -\frac{1}{3}x_4$ without indicating that one of these is a free variable. If we follow the convention of always stating basic variables in terms of free variables, and never the other way around, we avoid errors and unnecessary work.

Exercise 2.4.1. *For each system of equations, use an augmented matrix and row reduction to either find the solutions set or determine that the system is inconsistent.*

1. $$\begin{array}{rcrcrcrcr} x_1 & & & + & x_3 & & & = & 20 \\ & & x_2 & - & x_3 & - & x_4 & = & 0 \\ x_1 & + & x_2 & & & & & = & 80 \end{array}$$

2. $$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 4x_3 & = & 0 \\ 2x_1 & + & 3x_2 & + & 5x_3 & = & 0 \\ 3x_1 & + & 4x_2 & + & 2x_3 & = & 0 \end{array}$$

3. $$\begin{array}{rcrcrcr} 2x_1 & - & 2x_2 & + & x_3 & = & 6 \\ x_1 & + & x_2 & - & x_3 & = & -2 \\ & & x_2 & + & 3x_3 & = & 5 \end{array}$$

From the rref of the augmented matrix in Example 2.4.1, we see that the rightmost column, the augmented column, is a pivot column. This equates to the false equation "$0 = 1$," leading us to the conclusion that the system is inconsistent. What we see here is not unique to this example and in fact leads to a general relationship between the consistency of a linear system and the nature of the pivot columns of its augmented matrix. If the rightmost column of an augmented matrix is a pivot column, then the rref will include a row of the form

$$[0\ 0\ \cdots\ 0 \mid 1],$$

implying that the original system is equivalent to a system having the necessarily false equation "$0 = 1$." In fact, it is not necessary to obtain a full rref in order to conclude that a system is inconsistent. If any ref that is row equivalent to an augmented matrix includes a row of the form $[0\ 0\ \cdots\ 0 \mid \star]$, with $\star$ any nonzero number, the underlying system will include a false equation "$0 =$ something not zero."

We have seen that the pivot columns of an augmented matrix are related to the consistency of the underlying system, and they provide a way to characterize the variables and express solutions. We can state the following theorem that summarizes the relationship between pivot columns and the two big questions of existence and uniqueness.

Theorem 2.4.1. *Let A and $\hat{A}$ be the coefficient matrix and the augmented matrix of the linear system (2.3), respectively.*

1. *If the rightmost column of $\hat{A}$ is a pivot column of $\hat{A}$, then the system is inconsistent.*

2. *If the rightmost column of $\hat{A}$ is not a pivot column of $\hat{A}$, then the system is consistent.*

Moreover, if the system is consistent, then

1. *if every column of A is a pivot column of A, then the system has a unique solution; and*

2. *if at least one column of A is not a pivot column of A, then the system has infinitely many solutions.*

The first part of Theorem 2.4.1 reiterates our previous observation that a pivot position in the augmented column corresponds to a necessarily false equation ("$0 = 1$"). It also states that this must be the case if the system is inconsistent—meaning if the rightmost column of an augmented matrix is not a pivot column, the system must have at least one solution. As for the second part of Theorem 2.4.1, multiple solutions requires the presence of one or more free variables, and that requires at least one of the columns of A to not be a pivot column. Theorem 2.4.1 is summarized by the flow chart shown in Figure 2.2.

An immediate corollary to Theorem 2.4.1 is the following:

Corollary 2.4.1. *If $m < n$, then any system of m linear equations in n variables is either inconsistent or has infinitely many solutions.*

This says that if there are more variables than there are equations, it is not possible for the system to have a unique solution. It's certainly possible that such a system has no solution, but if it does have a solution, there must be at least one free variable (see Exercise 2.3.9).

Exercise 2.4.2. *For each augmented matrix, determine whether the associated system is inconsistent, consistent with a unique solution, or consistent with infinitely many solutions. (If possible, make your determinations without performing additional row operations.)*

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

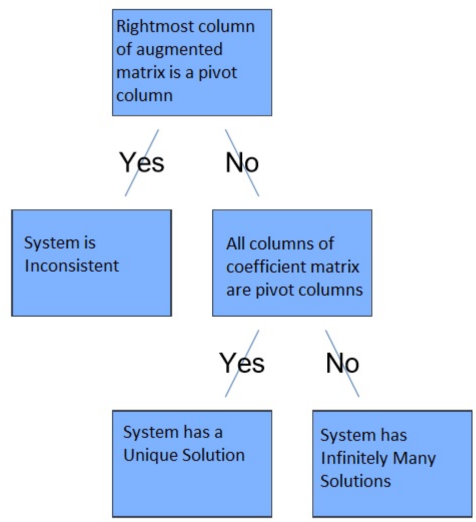


Figure 2.2: Flow Chart Illustrating Theorem 2.4.1

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

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

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

5. $\left[\begin{array}{cccccc|c} 1 & 2 & 0 & 0 & 0 & -2 & 4 \\ 0 & 0 & 1 & 0 & 0 & 4 & 2 \\ 0 & 0 & 0 & 1 & 0 & -5 & -2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right]$

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

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

8. $\left[\begin{array}{ccc|c} 1 & 0 & 0 & -8 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 0 \end{array}\right]$

Exercise 2.4.3. *For each of the consistent systems in Exercise 2.4.2, write the solution set in parametric form. Either assign parameters to any free variables, or be sure to clearly indicate which variables (if any) are free.(Note: you may need to perform additional row operations.)*

2.5 Additional Exercises

(Jump to Solutions)

1. Solve each linear system by using row reduction on the associated augmented matrix.

   a. $\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & x_3 & = & 1 \\ 3x_1 & + & 5x_2 & + & 3x_3 & = & 4 \\ 2x_1 & + & x_2 & + & x_3 & = & 4 \end{array}$

   b. $\begin{array}{rcrcrcr} x_1 & & & - & x_3 & = & 2 \\ 2x_1 & + & x_2 & + & 2x_3 & = & -6 \\ 3x_1 & + & 2x_2 & + & 2x_3 & = & -5 \end{array}$

   c. $\begin{array}{rcrcrcrcr} -2x_1 & + & 2x_2 & - & 3x_3 & - & 2x_4 & = & -8 \\ 3x_1 & - & 3x_2 & + & 3x_3 & + & x_4 & = & 10 \\ 2x_1 & - & 2x_2 & + & 2x_3 & & & = & 4 \end{array}$

   d. $\begin{array}{rcrcrcrcrcr} -2x_1 & - & 6x_2 & + & 4x_3 & - & 8x_4 & + & 32x_5 & = & 18 \\ 3x_1 & + & 9x_2 & + & x_3 & - & 2x_4 & - & 6x_5 & = & 8 \end{array}$

2. Determine all values of b, if any, such that the system of equations having the given augmented matrix is consistent.

   a. $\left[\begin{array}{cc|c} 2 & b & 3 \\ -1 & 3 & 4 \end{array}\right]$

b. $\left[\begin{array}{cc|c} 4 & 3 & -2 \\ 6 & 1 & b \end{array}\right]$

c. $\left[\begin{array}{cc|c} 4 & 6 & b \\ 6 & 9 & 12 \end{array}\right]$

3. For each system of equations, determine all value(s) of b and c, if any, such that the system of equations has (i) no solution, (ii) a unique solution, and (iii) infinitely many solutions.

   a. $$\begin{array}{rcrcl} x_1 & + & 3x_2 & = & 2 \\ 3x_1 & + & bx_2 & = & c \end{array}$$

   b. $$\begin{array}{rcrcl} bx_1 & - & 2x_2 & = & 5 \\ 4x_1 & + & 7x_2 & = & c \end{array}$$

   c. $$\begin{array}{rcrcl} 3x_1 & + & bx_2 & = & 0 \\ cx_1 & + & 4x_2 & = & 0 \end{array}$$

4. Create your own specific example of

   a. a system of linear equations with three equations and two variables that has a unique solution.

   b. a system of linear equations with three equations and two variables that is inconsistent.

   c. a system of linear equations with three equations and two variables that has infinitely many solutions.

   d. a linear equation with one variable that has a unique solution.

   e. a linear equation with one variable that is inconsistent.

   f. a linear equation with one variable that has infinitely many solutions.

5. Corollary 2.4.1 tells us that a system of linear equations that has more variables than equations either has no solution or has infinitely many solutions. (Such a system cannot have a unique solution.) Create your own specific example of

   a. a linear equation with two variables that has no solution.

   b. a linear equation with two variables that has infinitely many solutions.

c. a system of two linear equations with three variables that has no solution.

d. a system of two linear equations with three variables that has infinitely many solutions.

6. Find the solution set of the homogeneous system of linear equations having the given coefficient matrix.

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

b. $\begin{bmatrix} 2 & -1 & 7 \\ 3 & 1 & 3 \end{bmatrix}$

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

d. $\begin{bmatrix} 3 & 9 & 1 & -2 \\ 1 & 3 & -2 & 4 \end{bmatrix}$

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

7. Find an equation satisfied by g, h, and k such that the given matrix is the augmented matrix of a consistent linear system

$$\left[\begin{array}{rrr|c} 1 & -4 & 7 & g \\ 0 & 3 & -5 & h \\ -2 & 5 & -9 & k \end{array}\right]$$

8. Propane combines with oxygen to form carbon dioxide and water according to the chemical equation

$$x_1\ C_3H_8 + x_2\ O_2 \quad \longrightarrow \quad x_3\ CO_2 + x_4\ H_2O.$$

Balancing the number of atoms of carbon (C), hydrogen (H), and oxygen (O) leads to the homogeneous system of equations

$$\begin{array}{rcl} 3x_1 & = & x_3 \\ 8x_1 & = & 2x_4 \\ 2x_2 & = & 2x_3 + x_4 \end{array} .$$

Show that this system is homogeneous. Find the smallest positive integers x_1, x_2, x_3, x_4 that balance the chemical equation.

9. Boron sulfide and water react to produce boric acid and hydrogen sulfide gas according to the chemical equation

$$x_1 \text{ B}_2\text{S}_3 + x_2 \text{ H}_2\text{O} \quad \longrightarrow \quad x_3 \text{ H}_3\text{BO}_3 + x_4 \text{ H}_2\text{S}.$$

Balancing the number of atoms of boron (B), sulfer (S), hydrogen (H) and oxygen (O) leads to the homogeneous system of equations

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

Show that this system is homogeneous. Find the smallest positive integers x_1, x_2, x_3, x_4 that balance the chemical equation.

10. Suppose A is an $m \times n$ matrix whose i^{th} column is all zeros. Explain why the i^{th} column of rref(A) is all zeros.

11. Let

$$\vec{a}_1 = \langle 1, 0, 1, 0 \rangle, \ \ \vec{a}_2 = \langle -1, 2, 1, 1 \rangle, \ \ \vec{a}_3 = \langle 0, 0, 2, 2 \rangle, \ \text{ and } \ \vec{a}_4 = \langle 1, 1, 0, -1 \rangle.$$

Show that the vector $\vec{y} = \langle 2, -1, 3, 3 \rangle$ in R^4 is a linear combination of the vectors $\vec{a}_1, \vec{a}_2, \vec{a}_3$ and $\vec{a}_4$, and identify the weights. (Hint: the equation $x_1\vec{a}_1 + x_2\vec{a}_2 + x_3\vec{a}_3 + x_4\vec{a}_4 = \vec{y}$ can be translated into a linear system of equations for the weights $x_1, \ldots, x_4$.)

12. Determine whether the vector $\vec{x} = \langle -1, 3, 1 \rangle$ in R^3 is a linear combination of the vectors $\vec{u}$ and $\vec{v}$, where

$$\vec{u} = \langle 1, 1, -2 \rangle, \ \text{ and } \ \vec{v} = \langle 3, 2, 2 \rangle.$$

Chapter 3

Matrix Algebra

In Chapter 2, we learned that matrices are a useful tool for solving systems of linear equations. The solution set of any system of linear equations can be found by performing the row reduction algorithm on its augmented matrix. In the present chapter, we will see that matrices can be treated as algebraic objects in their own right. When we think of the term "algebra", we think of dealing with expressions and equations that probably contain some "constant" or "given" terms and might also contain some "variable" or "unknown" terms. We carry out "algebra" on such expressions or equations by performing a series of manipulations that involve legitimate operations that have been defined on the set of objects we are considering. The goal is often to find an unknown (or unknowns) in some problem.

We are already familiar with the operations of addition, subtraction, multiplication and division that are used in performing algebraic manipulations in problems involving real numbers. In Chapter 1, we defined operations of addition, subtraction, and scalar multiplication of vectors in R^n. In order to include matrices in the mix of objects that we can algebraically manipulate, we first need to define some operations on matrices. The operations that we will define for matrices will be

- addition and subtraction of two matrices
- multiplication of a matrix by a scalar
- multiplication of a matrix by a matrix
- transposition of a matrix

- multiplication of a vector by a matrix
- Inversion of a matrix

All of the above operations will be defined in such a way that they mesh properly with the operations that have already been defined for vectors in R^n. This will allow us to come up with concise formulations of problems that involve a combination of matrices, vectors, and scalars, and it will provide us with the algebraic tools that are needed to study these problems. Indeed, once we have defined matrix operations, we will see that systems of linear equations, which were studied in Chapter 2, can be formulated as matrix equations and analyzed using matrix algebra. But, as we will see as this linear algebra course unfolds, the usefulness of matrix algebra extends beyond solving systems of linear equations. In particular, regarding matrices as algebraic objects will allow us to use matrices to define functions that are called linear transformations between two Euclidean spaces R^n and R^m. Such functions are central to the subject of linear algebra.

As a prelude to our study of matrix algebra, we introduce some relevant notation that we will use throughout this chapter and throughout the remainder of the course.

3.1 Notation: Entries, Row Vectors and Column Vectors

In Chapter 2, we learned that a generic $m \times n$ matrix, A, can be written using the notation

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \tag{3.1}$$

whereby a matrix is named using an upper case letter and the entries of the matrix are named using the same lower case letter with subscripts. Using this naming convention, we can use the shorthand notation $A = [a_{ij}]$.

We will now introduce an alternative notation for identifying the entries of a matrix that will facilitate our study of matrix algebra. If A is an $m \times n$

matrix, then the entry in row i and column j of A will be denoted by $A_{(i,j)}$. For example, suppose that A is the matrix

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

The entry in row 2 and column 3 of A is 5. We can express this fact either by writing

$$a_{23} = 5 \quad \text{or} \quad A_{(2,3)} = 5.$$

At first, it may seem that the introduction of this new naming convention is unnecessary. However, the reason that we introduce it is that once we get down to the work of this chapter we will be wanting to refer to specific entries, not just of a single matrix, but of algebraic combinations of matrices such as products of matrices (to be defined in Section 3.3). Our newly introduced convention for identifying matrix entries will be more efficient in some situations. We will also still continue to use the original naming convention with lower case letters and subscripts when appropriate.

To carry out our work, we also need to define what is meant by the row vectors and column vectors of a matrix. For the generic $m \times n$ matrix, A, shown in (3.1), we define the **row vectors** of A to be the vectors

$$\begin{aligned} \text{Row}_1(A) &= \langle a_{11}, a_{12}, \ldots, a_{1n} \rangle \\ \text{Row}_2(A) &= \langle a_{21}, a_{22}, \ldots, a_{2n} \rangle \\ &\vdots \\ \text{Row}_m(A) &= \langle a_{m1}, a_{m2}, \ldots, a_{mn} \rangle \end{aligned}$$

and we define the **column vectors** of A to be the vectors

$$\begin{aligned} \text{Col}_1(A) &= \langle a_{11}, a_{21}, \ldots, a_{m1} \rangle \\ \text{Col}_2(A) &= \langle a_{12}, a_{22}, \ldots, a_{m2} \rangle \\ &\vdots \\ \text{Col}_n(A) &= \langle a_{1n}, a_{2n}, \ldots, a_{mn} \rangle . \end{aligned}$$

Note that row vectors of A are vectors in R^n and the column vectors of A are vectors in R^m.

As an example, suppose that A is the 3×4 matrix

$$A = \begin{bmatrix} -1 & -9 & 8 & 4 \\ -5 & 10 & -2 & 5 \\ 7 & 7 & -7 & 4 \end{bmatrix}.$$

The row vectors of A (which are vectors in R^4) are

$$\begin{aligned} \text{Row}_1(A) &= \langle -1, -9, 8, 4 \rangle \\ \text{Row}_2(A) &= \langle -5, 10, -2, 5 \rangle \\ \text{Row}_3(A) &= \langle 7, 7, -7, 4 \rangle \end{aligned}$$

and the column vectors of A (which are vectors in R^3) are

$$\begin{aligned} \text{Col}_2(A) &= \langle -1, -5, 7 \rangle \\ \text{Col}_2(A) &= \langle -9, 10, 7 \rangle \\ \text{Col}_3(A) &= \langle 8, -2, -7 \rangle \\ \text{Col}_4(A) &= \langle 4, 5, 4 \rangle \end{aligned}$$

Exercise 3.1.1. *For the* 5×4 *matrix*

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

express the entry in row 3 *and column* 3 *of* A *using two different notations (a notation involving the lower case* a *and a notation involving the upper case* A*). Do the same for the entry in row* 2 *and column* 4 *of* A.

Exercise 3.1.2. *1. Write down the row vectors and column vectors of the matrix*

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

Are the row vectors of A *in* R^4 *or in* R^2*? In what vector space do the column vectors of* A *live?*

2. *Write down the row vectors and column vectors of the matrix*

$$B = \begin{bmatrix} -5 & -2 & 1 \\ -8 & -2 & -6 \\ 5 & 5 & 3 \end{bmatrix}.$$

Where do the row vectors of B live? The column vectors?

3.2 Addition, Subtraction, and Scalar Multiplication of Matrices

The operations of addition, subtraction, and scalar multiplication of matrices are defined in a way that is analogous to how these operations were defined in Chapter 1 for vectors in R^n.

Specifically, if $A = [a_{ij}]$ and $B = [b_{ij}]$ are matrices of the same size, then we define $A + B$ to be the matrix

$$A + B = [a_{ij} + b_{ij}].$$

Likewise, we define

$$A - B = [a_{ij} - b_{ij}].$$

If $A = [a_{ij}]$ is a matrix and c is a scalar, then we define

$$cA = [ca_{ij}].$$

As examples of how we add and subtract matrices and multiply a matrix by a scalar, suppose that

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

and that $c = 3$. Then

$$A+B = \begin{bmatrix} -1+2 & 4+4 & -4+4 & 1+(-3) \\ -3+(-4) & 1+2 & 1+(-4) & -2+1 \\ 0+(-4) & 3+(-1) & 0+(-1) & -4+3 \end{bmatrix} = \begin{bmatrix} 1 & 8 & 0 & -2 \\ -7 & 3 & -3 & -1 \\ -4 & 2 & -1 & -1 \end{bmatrix}$$

and

$$A - B = \begin{bmatrix} -1-2 & 4-4 & -4-4 & 1-(-3) \\ -3-(-4) & 1-2 & 1-(-4) & -2-1 \\ 0-(-4) & 3-(-1) & 0-(-1) & -4-3 \end{bmatrix} = \begin{bmatrix} -3 & 0 & -8 & 4 \\ 1 & -1 & 5 & -3 \\ 4 & 4 & 1 & -7 \end{bmatrix}$$

and

$$cA = 3A = \begin{bmatrix} 3(-1) & 3(4) & 3(-4) & 3(1) \\ 3(-3) & 3(1) & 3(1) & 3(-2) \\ 3(0) & 3(3) & 3(0) & 3(-4) \end{bmatrix} = \begin{bmatrix} -3 & 12 & -12 & 3 \\ -9 & 3 & 3 & -6 \\ 0 & 9 & 0 & -12 \end{bmatrix}.$$

Exercise 3.2.1. *Suppose that A and B are the matrices*

$$A = \begin{bmatrix} -2 & 3 & -3 \\ 3 & 5 & 3 \\ 3 & 5 & -5 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 1 & -3 \\ 3 & -5 & 5 \end{bmatrix}$$

and suppose that $c = -2$ and $d = 2$. Perform the following computations.

1. $A + B$
2. $A - B$
3. $B - A$
4. cA
5. $cA + dB$

Exercise 3.2.2. *Suppose that A and B are the matrices*

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

and suppose that $c = 3$.

If possible, perform the following computations. If the computation you are being asked to perform is not possible, then explain why it is not possible.

1. $A + B$
2. $A - B$
3. cA
4. cB

3.2.1 Distributive Property

You are probably familiar with the **distributive property** of multiplication over addition for real numbers. This is the property that tells us that, when working with real numbers, multiplication "distributes" over addition. What this means is that if a, b, and c are any real numbers then $c\,(a+b) = ca + cb$. For example $3\,(4+7) = 3\,(11) = 33$ and $3\,(4) + 3\,(7) = 12 + 21 = 33$, so $3\,(4+7) = 3\,(4) + 3\,(7)$.

The distributive property also holds for distribution of scalar multiplication over matrix addition (or subtraction). If A and B are any two matrices of the same size and c is a scalar, then

$$c\,(A+B) = cA + cB.$$

The reason that the distributive property holds is that if $A = [a_{ij}]$ and $B = [b_{ij}]$, then

$$\begin{aligned} c\,(A+B) &= c\,([a_{ij}] + [b_{ij}]) \\ &= c\,[a_{ij} + b_{ij}] \\ &= [c\,(a_{ij} + b_{ij})] \\ &= [ca_{ij} + cb_{ij}] \\ &= [ca_{ij}] + [cb_{ij}] \\ &= c\,[a_{ij}] + c\,[b_{ij}] \\ &= cA + cB. \end{aligned}$$

As an illustration of the distributive property, suppose that $c = 3$ and

$$A = \begin{bmatrix} 4 & 8 \\ -6 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -1 & 4 \\ 1 & -3 \end{bmatrix}.$$

Then

$$\begin{aligned} c\,(A+B) &= 3\left(\begin{bmatrix} 4 & 8 \\ -6 & 1 \end{bmatrix} + \begin{bmatrix} -1 & 4 \\ 1 & -3 \end{bmatrix}\right) \\ &= 3\begin{bmatrix} 3 & 12 \\ -5 & -2 \end{bmatrix} \\ &= \begin{bmatrix} 9 & 36 \\ -15 & -6 \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} cA + cB &= 3\begin{bmatrix} 4 & 8 \\ -6 & 1 \end{bmatrix} + 3\begin{bmatrix} -1 & 4 \\ 1 & -3 \end{bmatrix} \\ &= \begin{bmatrix} 12 & 24 \\ -18 & 3 \end{bmatrix} + \begin{bmatrix} -3 & 12 \\ 3 & -9 \end{bmatrix} \\ &= \begin{bmatrix} 9 & 36 \\ -15 & -6 \end{bmatrix}, \end{aligned}$$

and we see that $c(A+B) = cA + cB$.

Exercise 3.2.3. *For the matrices A and B and scalars c given below, verify by computation that $c(A+B) = cA + cB$.*

1. $c = -5$ *and*

$$A = \begin{bmatrix} 1 & 0 \\ -2 & -3 \end{bmatrix} \quad \textit{and} \quad B = \begin{bmatrix} 1 & 1 \\ -3 & -3 \end{bmatrix}$$

2. $c = 3$ *and*

$$A = \begin{bmatrix} -2 & 3 & -1 \\ -1 & -3 & 2 \end{bmatrix} \quad \textit{and} \quad B = \begin{bmatrix} 3 & -1 & -2 \\ 1 & -3 & 2 \end{bmatrix}.$$

3.3 Multiplication of Two Matrices

We are going to define a way to multiply two matrices, A and B, to obtain another matrix. We will call this the **matrix product** AB. What might perhaps be your first guess on how we will make this definition (which would be the simplest way to make the definition) is to say that we can only form the product AB when $A = [a_{ij}]$ and $B = [b_{ij}]$ have the same size and we simply define $AB = [a_{ij}b_{ij}]$. Although this would indeed be simple, it would not be of any use in developing the subject of linear algebra. Instead, we are going to define the product AB only when A is an $m \times p$ matrix and B is a $p \times n$ matrix, meaning that the number of columns of A is the same as the number of rows of B. We will define AB in such a way that AB is an $m \times n$ matrix.

For example, if A is a 2×4 matrix and B is a 3×2 matrix, then AB *will not* be defined because A has 4 columns and B has 3 rows, so the number of columns of A does not match the number of rows of B. However BA *will* be defined because B has 2 columns and A has 2 rows, so the number of columns of B does match the number of rows of A. In this case, since B is a 3×2 matrix and A is a 2×4, then BA will be a 3×4 matrix.

Before defining the matrix product, we need to recall what is meant by the dot product of two vectors. In Section 1.3, we defined the dot product of two vectors $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle$ and $\vec{y} = \langle y_1, y_2, \ldots, y_n \rangle$ in R^n to be

$$\vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

For example if we have $\vec{x} = \langle 4, 0, 1 \rangle$ and $\vec{y} = \langle -1, -5, -7 \rangle$ (both of which are vectors in R^3), then

$$\vec{x} \cdot \vec{y} = (4)(-1) + (0)(-5) + (1)(-7) = -11.$$

And now for the definition of the matrix product: If A is an $m \times p$ matrix and B is a $p \times n$ matrix, then we define the **matrix product** AB to be the $m \times n$ matrix

$$AB = \left[\mathrm{Row}_i(A) \cdot \mathrm{Col}_j(B)\right]. \tag{3.2}$$

Thus AB is the $m \times n$ matrix whose entry in row i and column j is

$$(AB)_{(i,j)} = \mathrm{Row}_i(A) \cdot \mathrm{Col}_j(B)$$

Written more explicitly,

$$AB = \begin{bmatrix} \mathrm{Row}_1(A) \cdot \mathrm{Col}_1(B) & \mathrm{Row}_1(A) \cdot \mathrm{Col}_2(B) & \cdots & \mathrm{Row}_1(A) \cdot \mathrm{Col}_n(B) \\ \mathrm{Row}_2(A) \cdot \mathrm{Col}_1(B) & \mathrm{Row}_2(A) \cdot \mathrm{Col}_2(B) & \cdots & \mathrm{Row}_2(A) \cdot \mathrm{Col}_n(B) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Row}_m(A) \cdot \mathrm{Col}_1(B) & \mathrm{Row}_m(A) \cdot \mathrm{Col}_2(B) & \cdots & \mathrm{Row}_m(A) \cdot \mathrm{Col}_n(B) \end{bmatrix}.$$

For example, suppose that A and B are the matrices

$$A = \begin{bmatrix} -4 & -3 & -2 \\ 1 & -4 & -5 \\ 4 & -4 & 3 \\ 6 & 2 & -6 \end{bmatrix} \text{ and } B = \begin{bmatrix} 4 & -1 \\ 0 & 5 \\ 1 & 5 \end{bmatrix}.$$

Notice that A is a 4×3 matrix and B is a 3×2 matrix. Since the number of columns of A matches the number of rows of B $(3 = 3)$, then AB is defined and it is a 4×2 matrix. To find the entries of AB, we compute

$$\begin{aligned}
(AB)_{(1,1)} &= \text{Row}_1(A) \cdot \text{Col}_1(B) = \langle -4, -3, -2 \rangle \cdot \langle 4, 0, 1 \rangle = -18 \\
(AB)_{(1,2)} &= \text{Row}_1(A) \cdot \text{Col}_2(B) = \langle -4, -3, -2 \rangle \cdot \langle -1, 5, 5 \rangle = -21 \\
(AB)_{(2,1)} &= \text{Row}_2(A) \cdot \text{Col}_1(B) = \langle 1, -4, -5 \rangle \cdot \langle 4, 0, 1 \rangle = -1 \\
(AB)_{(2,2)} &= \text{Row}_2(A) \cdot \text{Col}_2(B) = \langle 1, -4, -5 \rangle \cdot \langle -1, 5, 5 \rangle = -46 \\
(AB)_{(3,1)} &= \text{Row}_3(A) \cdot \text{Col}_1(B) = \langle 4, -4, 3 \rangle \cdot \langle 4, 0, 1 \rangle = 19 \\
(AB)_{(3,2)} &= \text{Row}_3(A) \cdot \text{Col}_2(B) = \langle 4, -4, 3 \rangle \cdot \langle -1, 5, 5 \rangle = -9 \\
(AB)_{(4,1)} &= \text{Row}_4(A) \cdot \text{Col}_1(B) = \langle 6, 2, -6 \rangle \cdot \langle 4, 0, 1 \rangle = 18 \\
(AB)_{(4,2)} &= \text{Row}_4(A) \cdot \text{Col}_2(B) = \langle 6, 2, -6 \rangle \cdot \langle -1, 5, 5 \rangle = -26.
\end{aligned}$$

We conclude that

$$AB = \begin{bmatrix} -18 & -21 \\ -1 & -46 \\ 19 & -9 \\ 18 & -26 \end{bmatrix}.$$

Notice that it is not possible to compute the matrix product BA for the above two matrices. The reason is that B is a 3×2 matrix and A is a 4×3 matrix, and thus the number of columns of B (which is 2) does not match the number of rows of A (which is 4). Thus BA is undefined. This example alerts us to an important issue regarding matrix multiplication, which is that matrix multiplication is **not commutative**. When we multiply two real numbers, say 2 and 5, the order in which we multiply them does not matter: $(2)(5) = 10$ and $(5)(2) = 10$. We call this the **commutative property of multiplication** of real numbers. However, the order in which we multiply matrices *does* matter. For the matrices A and B in the above example, we certainly cannot say that $AB = BA$ because BA is not even defined. In fact, even if AB and BA are both defined, then it is typically still not true that $AB = BA$. It is possible that AB and BA are both defined but are of different sizes. For example, if A has size 3×4 and B has size 4×3, then AB and BA are both defined, but AB has size 3×3 and BA has size 4×4, so obviously $AB \neq BA$. In order for AB and BA to both be defined and have the same size, A and B must both be square matrices of the same size. A **square matrix** is a matrix that has the same number of rows as columns. So for example, if A has size 3×3 and B has size 3×3, then A and B are both square matrices and they have the same size. AB and BA are both defined and each of them has size 3×3. However, even in this case, it is *still* usually not true that $AB = BA$. For example, for the pair of 2×2 matrices

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

we have

$$AB = \begin{bmatrix} 6 & -12 \\ 2 & -12 \end{bmatrix} \text{ and } BA = \begin{bmatrix} -8 & -4 \\ -8 & 2 \end{bmatrix}$$

and thus $AB \neq BA$.

Although it is usually not true that $AB = BA$, we should point out that it is true for some matrices A and B. For example, the matrices

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

satisfy $AB = BA$ (as the reader should verify).

Exercise 3.3.1. *For each of the following pairs of matrices, A and B, compute both AB and BA (assuming they are defined). Then state whether $AB = BA$ or $AB \neq BA$.*

1.

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

2.

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

3.

$$A = \begin{bmatrix} 2 & 5 \\ 0 & 7 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & -4 & -7 \\ -7 & 0 & -1 \end{bmatrix}$$

4.

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 0 & -1 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 & 0 \\ 0 & -3 \\ -1 & 2 \end{bmatrix}$$

5.

$$A = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} -2 & 3 & -1 \end{bmatrix}$$

6.

$$A = \begin{bmatrix} 2 & 3 \\ -4 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 & 3 \\ -4 & 0 \end{bmatrix}$$

7.

$$A = \begin{bmatrix} 1 & 3 \\ 5 & -4 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

8.

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} -1 & 2 \\ 4 & -1 \end{bmatrix}$$

9.

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -12 & 0 \\ 0 & 0 & 4 \end{bmatrix} \text{ and } B = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

10.

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 5 \\ -2 & -5 \end{bmatrix}.$$

Although matrix multiplication is not commutative, there is a frequently used property of matrix multiplication that tells us that a scalar that appears in a product involving two matrices and that scalar can "commute". Specifically, if A and B are two matrices such that the matrix product AB is defined (i.e., the number of columns of A is the same as the number of rows of B) and c is a scalar, then

$$c\,(AB) = (cA)\,B = A\,(cB)\,. \tag{3.3}$$

Example 3.3.1. *We will illustrate the property (3.3) for the scalar $c = 3$ and the matrices*

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

First we compute

$$\begin{aligned} 3\,(AB) &= 3\left(\begin{bmatrix} -1 & 2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 4 & 1 \end{bmatrix}\right) \\ &= 3\begin{bmatrix} 9 & 1 \\ 4 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 27 & 3 \\ 12 & 3 \end{bmatrix}. \end{aligned}$$

Next we compute

$$\begin{aligned} (3A)\,B &= \left(3\begin{bmatrix} -1 & 2 \\ 0 & 1 \end{bmatrix}\right)\begin{bmatrix} -1 & 1 \\ 4 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -3 & 6 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 4 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 27 & 3 \\ 12 & 3 \end{bmatrix}. \end{aligned}$$

Finally we compute

$$\begin{aligned} A\,(3B) &= \begin{bmatrix} -1 & 2 \\ 0 & 1 \end{bmatrix}\left(3\begin{bmatrix} -1 & 1 \\ 4 & 1 \end{bmatrix}\right) \\ &= \begin{bmatrix} -1 & 2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -3 & 3 \\ 12 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 27 & 3 \\ 12 & 3 \end{bmatrix}. \end{aligned}$$

As can be seen,

$$3\,(AB) = (3A)\,B = A\,(3B) = \begin{bmatrix} 27 & 3 \\ 12 & 3 \end{bmatrix}.$$

Exercise 3.3.2. *Verify, by computation, that property (3.3) holds for the following scalars and pairs of matrices.*

1. $c = 4$ *and*

$$A = \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix} \text{ and } B = \begin{bmatrix} -3 & 1 \\ 1 & -3 \end{bmatrix}$$

2. $c = -2$ *and*

$$A = \begin{bmatrix} -3 & -2 \\ 3 & -1 \end{bmatrix} \text{ and } B = \begin{bmatrix} -1 & -3 & 2 \\ 3 & -1 & -2 \end{bmatrix}$$

3. $c = 2$ *and*

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

As a word of caution, remember that matrix multiplication is not commutative! This means that, in applying the property (3.3), we **may not** switch the order of the matrices A and B. For example, it is usually **not** true that $(cA)\,B = (cB)\,A$.

3.3.1 Distributive Property

Just as there is a distributive property of scalar multiplication over matrix addition, there is also a distributive property of matrix multiplication over matrix addition. If B and C are two matrices of the same size, and A is a matrix such that the number of columns of A is the same as the number of rows of B (and hence the same as the number of rows of C), then

$$A\,(B + C) = AB + AC. \tag{3.4}$$

To see why this distributive property holds, we need to use our definition of matrix multiplication (3.2) along with the fact that the dot product is distributive over vector addition. The distributive property of the dot product over vector addition says that if $\vec{x}$, $\vec{y}$, and $\vec{z}$ are any three vectors in R^n, then

$$\vec{x} \cdot (\vec{y} + \vec{z}) = \vec{x} \cdot \vec{y} + \vec{x} \cdot \vec{z}.$$

You were asked to verify this distributive property in Exercise 1.1.17

Now, to verify the distributive property of matrix multiplication over matrix addition, suppose that B and C are two matrices of the same size, and that A is a matrix such that the number of columns of A is the same as the number of rows of B (and hence the same as the number of rows of C). This implies that $A\,(B + C)$ and $AB + AC$ are both defined and both have the same size. To see why they are actually the same matrix, let us pick a specific (but arbitrary) row and column of $A(B+C)$ and show that the corresponding entry of $AB+AC$ is the same. By the definition of matrix multiplication (3.2), the entry in row i and column j of $A(B+C)$ is

$$\mathrm{Row}_i(A)\cdot\mathrm{Col}_j(B+C)=\mathrm{Row}_i(A)\cdot(\mathrm{Col}_j(B)+\mathrm{Col}_j(C)).$$

Using the distributive property of the dot product, we see that

$$\mathrm{Row}_i(A)\cdot(\mathrm{Col}_j(B)+\mathrm{Col}_j(C))=\mathrm{Row}_i(A)\cdot\mathrm{Col}_j(B)+\mathrm{Row}_i(A)\cdot\mathrm{Col}_j(C).$$

However $\mathrm{Row}_i(A)\cdot\mathrm{Col}_j(B)$ is the entry in row i and column j of AB and $\mathrm{Row}_i(A)\cdot\mathrm{Col}_j(C)$ is the entry in row i and column j of AC. This implies that $\mathrm{Row}_i(A)\cdot\mathrm{Col}_j(B)+\mathrm{Row}_i(A)\cdot\mathrm{Col}_j(C)$ is the entry in row i and column j of $AB+AC$.

Exercise 3.3.3. *For the matrices A, B and C given below, verify by computation that $A(B+C)=AB+AC$.*

1.
$$A=\begin{bmatrix}3&0\\1&3\end{bmatrix},\qquad B=\begin{bmatrix}1&5\\2&0\end{bmatrix},\qquad C=\begin{bmatrix}-1&1\\0&-2\end{bmatrix}.$$

2.
$$A=\begin{bmatrix}-3&1\\3&-3\end{bmatrix},\quad B=\begin{bmatrix}-3&-2&3\\-1&-1&-3\end{bmatrix},\quad C=\begin{bmatrix}0&3&-1\\-2&1&-3\end{bmatrix}$$

Exercise 3.3.4. *Suppose A and B are $m\times p$ matrices and C is a $p\times n$ matrix. Use the definition of matrix multiplication along with the distributive property $(\vec{x}+\vec{y})\cdot\vec{z}=\vec{x}\cdot\vec{z}+\vec{y}\cdot\vec{z}$ to verify that*

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

Exercise 3.3.5. *For the matrices A, B and C given below, verify by computation that $(A+B)C=AC+BC$.*

$$A=\begin{bmatrix}3&0\\1&3\end{bmatrix},\qquad B=\begin{bmatrix}1&5\\2&0\end{bmatrix},\qquad C=\begin{bmatrix}-1&1\\0&-2\end{bmatrix}.$$

3.4 The Transpose of a Matrix

A matrix that is closely related to any given $m \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

is the **transpose** of A, which is the matrix

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \vdots & a_{m1} \\ a_{12} & a_{22} & \vdots & a_{m2} \\ \cdots & \cdots & \ddots & \cdots \\ a_{1n} & a_{2n} & \vdots & a_{mn} \end{bmatrix}.$$

Thus A^T (which is pronounced as "A transpose") is the matrix such that $\text{Col}_i\left(A^T\right) = \text{Row}_i\left(A\right)$ for all $i = 1, 2, \ldots, m$. Another way to look at it is that $\text{Row}_i\left(A^T\right) = \text{Col}_i\left(A\right)$ for all $i = 1, 2, \ldots, n$. Yet another way to look at it as that $\left[A^T\right]_{(i,j)} = [A]_{(j,i)}$.

As an example, suppose that A is the matrix

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

Then

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

Notice that, in the above example, we have

$$\left(A^T\right)^T = \begin{bmatrix} -1 & -1 & 0 \\ 2 & -1 & 1 \end{bmatrix} = A.$$

This is a general property of transposes. If A is any matrix, then

$$\left(A^T\right)^T = A.$$

It is also true (and not difficult to prove) that if A is any matrix and c is any scalar, then

$$(cA)^T = c\left(A^T\right).$$

Another useful property of transposes is that if A and B are matrices of appropriate sizes such that the matrix product AB is defined, then

$$(AB)^T = B^T A^T. \tag{3.5}$$

In words, (3.5) says that the transpose of the product of two matrices is equal to the product of the transposes in the reverse order. It is important that we pay attention to the *in the reverse order* part of this statement because matrix multiplication is not commutative. It is *not* generally true that $(AB)^T = A^T B^T$.

To see why property (3.5) holds, first note that if A is an $m \times p$ matrix and B is a $p \times n$ matrix, then AB is an $m \times n$ matrix, which means that $(AB)^T$ is an $n \times m$ matrix. Also, B^T is an $n \times p$ matrix and A^T is a $p \times m$ matrix, which means that $B^T A^T$ is an $n \times m$ matrix. Therefore $(AB)^T$ and $B^T A^T$ have the same size. To see why these two matrices are equal to each other, note that the entry in row i and column j of $(AB)^T$ is

$$\left[(AB)^T\right]_{(i,j)} = [AB]_{(j,i)} = \text{Row}_j(A) \cdot \text{Col}_i(B)$$

and that the entry in row i and column j of $B^T A^T$ is

$$\left[B^T A^T\right]_{(i,j)} = \text{Row}_i\left(B^T\right) \cdot \text{Col}_j\left(A^T\right) = \text{Col}_i(B) \cdot \text{Row}_j(A)$$

and these are the same (because the dot product is commutative).

Let us use the matrices

$$A = \begin{bmatrix} -1 & -1 \\ 2 & -1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

to illustrate property (3.5). We see that

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

and thus

$$(AB)^T = \begin{bmatrix} -3 & 3 \\ 1 & -2 \\ -1 & -1 \end{bmatrix}.$$

Now observe that

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

and thus

$$B^T A^T = \begin{bmatrix} 2 & 1 \\ -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} -3 & 3 \\ 1 & -2 \\ -1 & -1 \end{bmatrix}.$$

We see that $(AB)^T = B^T A^T$.

Exercise 3.4.1. *Suppose that matrix A has size 5×7 and that matrix B has size 7×3.*

1. *What is the size of A^T?*
2. *What is the size of B^T?*
3. *What is the size of $(AB)^T$?*

Exercise 3.4.2. *For each of the matrices, A, given below, find A^T. Then find $\left(A^T\right)^T$ and observe that it is equal to A.*

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

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

Exercise 3.4.3. *For each of the matrices, A, and scalars, c, given below, verify that $(cA)^T = c\left(A^T\right)$.*

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

2.

$$A = \begin{bmatrix} 4 & 2 \\ -1 & 0 \end{bmatrix} \text{ and } c = -5$$

Exercise 3.4.4. *For each pair of matrices, A and B, given below, verify that* $(AB)^T = B^T A^T$.

1.

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

2.

$$A = \begin{bmatrix} 0 & 1 & 2 \\ -2 & 0 & 0 \\ 2 & 0 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 5 & -2 \\ 1 & 1 \\ -1 & 0 \end{bmatrix}.$$

Exercise 3.4.5. *Only the first row of the matrix*

$$A = \begin{bmatrix} 1 & 2 & -6 \\ \text{--} & \text{--} & \text{--} \\ \text{--} & \text{--} & \text{--} \end{bmatrix},$$

is given. Is it possible to fill in the remaining two rows of A in such a way that the statement $A^T = A$ *is true? If so, then do it. If not, then explain why not.*

Only three entries of the matrix

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

are given. Is it possible to fill in the remaining entries of A in such a way that the statement $A^T = A$ *is true? If so, then do it. If not, then explain why not.*

3.5 Multiplication of a Vector by a Matrix

Next on our agenda is to define multiplication of a vector $\vec{x}$ by a matrix A. This matrix–vector product will be denoted by $A\vec{x}$ and will be a vector. The definition will be made so as to be compatible with our definition of matrix multiplication (3.2).

Suppose that A is a matrix of of size $m \times n$ and that $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle$ is a vector in R^n. We define the **matrix–vector product** $A\vec{x}$ to be the vector

$$A\vec{x} = \langle \text{Row}_1(A) \cdot \vec{x}, \text{Row}_2(A) \cdot \vec{x}, \ldots, \text{Row}_m(A) \cdot \vec{x} \rangle . \tag{3.6}$$

The reason that this definition is compatible with the definition of matrix multiplication (3.2) is that if we form the matrix

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

that consists of a single column whose entries are the entries of $\vec{x}$, then by definition (3.2) we obtain

$$AX = \begin{bmatrix} \text{Row}_1(A) \cdot \text{Col}_1(X) \\ \text{Row}_2(A) \cdot \text{Col}_1(X) \\ \vdots \\ \text{Row}_m(A) \cdot \text{Col}_1(X) \end{bmatrix},$$

but since $\text{Col}_1(X) = \vec{x}$, then we have

$$AX = \begin{bmatrix} \text{Row}_1(A) \cdot \vec{x} \\ \text{Row}_2(A) \cdot \vec{x} \\ \vdots \\ \text{Row}_m(A) \cdot \vec{x} \end{bmatrix}.$$

Notice that if A is an $m \times n$ matrix and $\vec{x}$ is a vector in R^n, then $A\vec{x}$ is a vector in R^m.

Example 3.5.1. *Let us compute $A\vec{x}$ for the 3×4 matrix*

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

and $\vec{x} = \langle 1, -3, 0, 2 \rangle$ in R^4.

Using the definition (3.6), we obtain

$$\begin{aligned}\mathrm{Row}_1(A)\cdot\vec{x} &= \langle 3,0,1,3\rangle\cdot\langle 1,-3,0,2\rangle = 9\\ \mathrm{Row}_2(A)\cdot\vec{x} &= \langle 1,-1,2,0\rangle\cdot\langle 1,-3,0,2\rangle = 4\\ \mathrm{Row}_3(A)\cdot\vec{x} &= \langle 0,2,0,-1\rangle\cdot\langle 1,-3,0,2\rangle = -8\end{aligned}$$

and thus $A\vec{x} = \langle 9,4,-8\rangle$

There is an alternative way to view the matrix product $A\vec{x}$ for a given $m\times n$ matrix $A=[a_{ij}]$ and a given vector $\vec{x}=\langle x_1,x_2,\ldots,x_n\rangle$ in R^n. This alternative view is that

$$A\vec{x} = x_1\,\mathrm{Col}_1(A) + x_2\,\mathrm{Col}_2(A) + \cdots + x_n\,\mathrm{Col}_n(A). \tag{3.7}$$

To see why (3.6) and (3.7) are equivalent, note that the ith component of the vector $x_1\,\mathrm{Col}_1(A)$ is x_1a_{i1}, the ith component of the vector $x_2\,\mathrm{Col}_2(A)$ is x_2a_{i2}, etc. Thus the ith component of the vector $A\vec{x}$ as defined by (3.7) is

$$x_1a_{i1} + x_2a_{i2} + \cdots + x_na_{in}.$$

On the other hand, if we use (3.6) to compute $A\vec{x}$, then the ith component of $A\vec{x}$ is

$$\mathrm{Row}_i(A)\cdot\vec{x} = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n$$

which is the same thing we get when we use (3.7) to compute $A\vec{x}$.

In writing $A\vec{x}$ in the form (3.7), we are expressing $A\vec{x}$ as a linear combination of the column vectors of A using the entries of $\vec{x}$ as weights. (The reader may want to refer back to Definition 1.3.1 to review the concept of linear combinations of vectors).

Example 3.5.2. *In Example 3.5.1, we computed* $A\vec{x}$ *for the matrix*

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

and the vector $\vec{x}=\langle 1,-3,0,2\rangle$*. We did that computation using (3.6), which involves computing dot products of each of the row vectors of* A *with* $\vec{x}$*. Let us now see how the computation works if we instead use (3.7). We have*

$$\begin{aligned}A\vec{x} &= x_1\,\mathrm{Col}_1(A) + x_2\,\mathrm{Col}_2(A) + x_3\,\mathrm{Col}_3(A) + x_4\,\mathrm{Col}_4(A)\\ &= (1)\langle 3,1,0\rangle + (-3)\langle 0,-1,2\rangle + (0)\langle 1,2,0\rangle + (2)\langle 3,0,-1\rangle\\ &= \langle 3,1,0\rangle + \langle 0,3,-6\rangle + \langle 0,0,0\rangle + \langle 6,0,-2\rangle\\ &= \langle 9,4,-8\rangle.\end{aligned}$$

It is interesting to observe that one of the two ways to compute $A\vec{x}$ involves the row vectors of A and the other way to compute $A\vec{x}$ involves the column vectors of A. It is worthwhile to understand both approaches, so we now provide some exercises that will give you practice.

Exercise 3.5.1. *For each of the matrices, A, and vectors, $\vec{x}$, given in 1–6 below, compute $A\vec{x}$ in two different ways:* ***a)*** *by using (3.6) and* ***b)*** *by using (3.7).*

1. $$A = \begin{bmatrix} -1 & -2 \\ 0 & 1 \end{bmatrix}, \qquad \vec{x} = \langle 3, 5 \rangle$$

2. $$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & -1 & -1 \end{bmatrix}, \qquad \vec{x} = \langle 5, -1, 0 \rangle$$

3. $$A = \begin{bmatrix} -2 & 2 & 5 & -1 \\ 4 & 0 & 2 & -1 \\ 1 & -1 & 2 & 1 \end{bmatrix}, \qquad \vec{x} = \langle 2, -3, -3, 5 \rangle$$

4. $$A = \begin{bmatrix} -2 & 1 \\ -3 & 2 \\ 3 & -2 \\ -2 & 0 \end{bmatrix}, \qquad \vec{x} = \langle -4, -2 \rangle$$

5. $$A = \begin{bmatrix} 3 & 0 & 1 \\ 3 & 1 & -1 \\ 2 & 0 & -1 \end{bmatrix}, \qquad \vec{x} = \langle x_1, x_2 . x_3 \rangle$$

6. $$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix}, \qquad \vec{x} = \langle x_1, x_2, x_3, x_4 \rangle$$

7. *Write down the matrix A and the vector $\vec{x}$ such that $A\vec{x} =$*

$$\langle \langle 6, -3, -2 \rangle \cdot \langle 1, -3, 1 \rangle, \langle 5, -2, 2 \rangle \cdot \langle 1, -3, 1 \rangle, \langle -3, -3, -1 \rangle \cdot \langle 1, -3, 1 \rangle \rangle .$$

8. *Write down the matrix A and the vector $\vec{x}$ such that*

$$A\vec{x} = (7)\langle 0,-1,-6\rangle + (-7)\langle -4,-5,-1\rangle + (4)\langle 3,-7,4\rangle .$$

9. *For the matrix A and the vector $\vec{x}$ that you found in problem 7, write $A\vec{x}$ as a linear combination of the column vectors of A.*

10. *For the matrix A and the vector $\vec{x}$ that you found in problem 8, write $A\vec{x}$ using dot products of the row vectors of A with $\vec{x}$.*

Let us now look at how multiplication of a vector by a matrix is related to the multiplication of two matrices. Recall that we have defined the product (3.2) of the $m \times p$ matrix A and the $p \times n$ matrix B to be the $m \times n$ matrix

$$AB = \begin{bmatrix} \mathrm{Row}_1(A)\cdot\mathrm{Col}_1(B) & \mathrm{Row}_1(A)\cdot\mathrm{Col}_2(B) & \cdots & \mathrm{Row}_1(A)\cdot\mathrm{Col}_n(B) \\ \mathrm{Row}_2(A)\cdot\mathrm{Col}_1(B) & \mathrm{Row}_2(A)\cdot\mathrm{Col}_2(B) & \cdots & \mathrm{Row}_2(A)\cdot\mathrm{Col}_n(B) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Row}_m(A)\cdot\mathrm{Col}_1(B) & \mathrm{Row}_m(A)\cdot\mathrm{Col}_2(B) & \cdots & \mathrm{Row}_m(A)\cdot\mathrm{Col}_n(B) \end{bmatrix}.$$

The observation that we wish to make is that for any given column of AB (say the ith column) we have

$$\mathrm{Col}_i(AB) = A\,\mathrm{Col}_i(B). \tag{3.8}$$

Thus the ith column vector of the matrix AB is the same as the matrix A multiplied by the vector $\mathrm{Col}_i(B)$.

Example 3.5.3. *We will illustrate the property (3.8) for the matrices*

$$A = \begin{bmatrix} 1 & -4 \\ 1 & 4 \end{bmatrix} \quad \textit{and} \quad B = \begin{bmatrix} -4 & 2 \\ -2 & -4 \end{bmatrix}.$$

First we compute

$$\begin{aligned} \mathrm{Row}_1(A)\cdot\mathrm{Col}_1(B) &= \langle 1,-4\rangle\cdot\langle -4,-2\rangle = 4 \\ \mathrm{Row}_2(A)\cdot\mathrm{Col}_1(B) &= \langle 1,4\rangle\cdot\langle -4,-2\rangle = -12 \\ \mathrm{Row}_1(A)\cdot\mathrm{Col}_2(B) &= \langle 1,-4\rangle\cdot\langle 2,-4\rangle = 18 \\ \mathrm{Row}_2(A)\cdot\mathrm{Col}_2(B) &= \langle 1,4\rangle\cdot\langle 2,-4\rangle = -14 \end{aligned}$$

and then observe that

$$A\,\mathrm{Col}_1(B) = \langle \mathrm{Row}_1(A)\cdot \mathrm{Col}_1(B), \mathrm{Row}_2(A)\cdot \mathrm{Col}_1(B)\rangle = \langle 4, -12\rangle$$

and

$$A\,\mathrm{Col}_2(B) = \langle \mathrm{Row}_1(A)\cdot \mathrm{Col}_2(B), \mathrm{Row}_2(A)\cdot \mathrm{Col}_2(B)\rangle = \langle 18, -14\rangle .$$

Also, by direct computation

$$AB = \begin{bmatrix} 4 & 18 \\ -12 & -14 \end{bmatrix}.$$

We observe that

$$\mathrm{Col}_1(AB) = A\,\mathrm{Col}_1(B) = \langle 4, -12\rangle$$

and

$$\mathrm{Col}_2(AB) = A\,\mathrm{Col}_2(B) = \langle 18, -14\rangle .$$

Exercise 3.5.2. *Let*

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

and

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

Suppose that we are only interested in computing $\mathrm{Col}_2(AB)$*. Since* A *and* B *are rather large matrices, it would be a lot of work to compute* $\mathrm{Col}_2(AB)$ *by actually first computing* AB *itself and then looking to see what the second column of* AB *is. However, it is not as much work if we use the fact (3.8). Use this fact to compute* $\mathrm{Col}_2(AB)$*.*

Exercise 3.5.3. *Suppose that* A *is a* 4×5 *matrix and that* B *is a* 5×3 *matrix and that the second column of* B *consists entirely of entries of* 0*.*

1. *What size is the matrix AB?*
2. *What can you say about the second column of AB?*

We will conclude this section by looking at how multiplication of a vector by a matrix is related to multiplication of a vector by the transpose of that matrix. Suppose that A is an $m \times n$ matrix and recall from Section 3.4 that A^T is the $n \times m$ matrix for which

$$\begin{aligned} \mathrm{Row}_i\left(A^T\right) &= \mathrm{Col}_i\left(A\right) \\ \mathrm{Col}_i\left(A^T\right) &= \mathrm{Row}_i\left(A\right) \\ \left(A^T\right)_{(i,j)} &= A_{(j,i)} \end{aligned}$$

for all relevant indices i and j.

By definition (3.6), we see that if $\vec{x}$ is a vector in R^m, then

$$A^T\vec{x} = \left\langle \mathrm{Row}_1\left(A^T\right)\cdot\vec{x}, \mathrm{Row}_2\left(A^T\right)\cdot\vec{x}, \ldots, \mathrm{Row}_n\left(A\right)\cdot\vec{x}\right\rangle.$$

Since $\mathrm{Row}_i\left(A^T\right) = \mathrm{Col}_i\left(A\right)$ for all $i = 1, 2, \ldots, n$, then

$$A^T\vec{x} = \left\langle \mathrm{Col}_1\left(A\right)\cdot\vec{x}, \mathrm{Col}_2\left(A\right)\cdot\vec{x}, \ldots, \mathrm{Col}_n\left(A\right)\cdot\vec{x}\right\rangle \tag{3.9}$$

Likewise, by (3.7), we have

$$A^T\vec{x} = x_1\,\mathrm{Col}_1\left(A^T\right) + x_2\,\mathrm{Col}_2\left(A^T\right) + \cdots + x_m\,\mathrm{Col}_m\left(A^T\right).$$

Since $\mathrm{Col}_i\left(A^T\right) = \mathrm{Row}_i\left(A\right)$ for all $i = 1, 2, \ldots, m$ then

$$A^T\vec{x} = x_1\,\mathrm{Row}_1\left(A\right) + x_2\,\mathrm{Row}_2\left(A\right) + \cdots + x_m\,\mathrm{Row}_m\left(A\right) \tag{3.10}$$

Observe that (3.10) expresses $A^T\vec{x}$ as a linear combination of the row vectors of A using the entries of $\vec{x}$ as weights.

Example 3.5.4. *Suppose that A is the matrix*

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

and that $\vec{x} = \langle 3, 2, 3\rangle$. Let us use (3.9) to compute $A^T\vec{x}$.

We have

$$\begin{aligned} \mathrm{Col}_1\left(A\right)\cdot\vec{x} &= \langle -1, 0, -1\rangle\cdot\langle 3, 2, 3\rangle = -6 \\ \mathrm{Col}_2\left(A\right)\cdot\vec{x} &= \langle -2, 1, 1\rangle\cdot\langle 3, 2, 3\rangle = -1 \end{aligned}$$

and thus $A^T\vec{x} = \langle -6, -1\rangle$.

Exercise 3.5.4. *For each of the matrices, A, and vectors, $\vec{x}$, given in 1–6 below, compute $A^T\vec{x}$ in two different ways:* ***a)*** *by using (3.9) and* ***b)*** *by using (3.10).*

1. $$A = \begin{bmatrix} -1 & -2 \\ 0 & 1 \end{bmatrix}, \qquad \vec{x} = \langle 3, 5 \rangle$$

2. $$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & -1 & -1 \end{bmatrix}, \qquad \vec{x} = \langle 5, -1, 0 \rangle$$

3. $$A = \begin{bmatrix} -2 & 2 & 5 & -1 \\ 4 & 0 & 2 & -1 \\ 1 & -1 & 2 & 1 \end{bmatrix}, \qquad \vec{x} = \langle -3, 3, -3 \rangle$$

4. $$A = \begin{bmatrix} -2 & 1 \\ -3 & 2 \\ 3 & -2 \\ -2 & 0 \end{bmatrix}, \qquad \vec{x} = \langle 3, 3, -2, -3 \rangle$$

5. $$A = \begin{bmatrix} 3 & 0 & 1 \\ 3 & 1 & -1 \\ 2 & 0 & -1 \end{bmatrix}, \qquad \vec{x} = \langle x_1, x_2 . x_3 \rangle$$

6. $$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix}, \qquad \vec{x} = \langle x_1, x_2 \rangle$$

7. *Write down the matrix A and the vector $\vec{x}$ such that*
$$A^T\vec{x} = \langle \langle -2, -2 \rangle \cdot \langle 1, -1 \rangle, \langle 1, -2 \rangle \cdot \langle 1, -1 \rangle, \langle -1, 3 \rangle \cdot \langle 1, -1 \rangle \rangle .$$

8. *Write down the matrix A and the vector $\vec{x}$ such that*
$$A^T\vec{x} = (-3) \langle -4, -1 \rangle + (-7) \langle 1, 0 \rangle + (2) \langle 2, -2 \rangle .$$

9. *For the matrix A and the vector $\vec{x}$ that you found in problem 7, write $A^T\vec{x}$ as a linear combination of the row vectors of A.*

10. *For the matrix A and the vector $\vec{x}$ that you found in problem 8, write $A^T\vec{x}$ using dot products of the column vectors of A with $\vec{x}$.*

We have seen that if A is an $m \times p$ matrix and B is a $p \times n$ matrix then

$$\mathrm{Col}_i\left(AB\right) = A\,\mathrm{Col}_i\left(B\right).$$

An analogous fact is that

$$\mathrm{Row}_i\left(AB\right) = B^T\,\mathrm{Row}_i\left(A\right). \tag{3.11}$$

To see why (3.11) is true, observe that for any $i = 1, 2, \ldots, m$, we have

$$\begin{aligned}
\mathrm{Row}_i\left(AB\right) &= \mathrm{Col}_i\left(\left(AB\right)^T\right) \\
&= \mathrm{Col}_i\left(B^T A^T\right) \\
&= B^T\,\mathrm{Col}_i\left(A^T\right) \\
&= B^T\,\mathrm{Row}_i\left(A\right)
\end{aligned}$$

Exercise 3.5.5. *Illustrate the property (3.11) for the matrices*

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

In other words, show by computation that

$$\mathrm{Row}_1\left(AB\right) = B^T\,\mathrm{Row}_1\left(A\right)$$

and

$$\mathrm{Row}_2\left(AB\right) = B^T\,\mathrm{Row}_2\left(A\right).$$

Exercise 3.5.6. *Suppose that A is a 12×23 matrix and that B is a 23×5 matrix. Suppose, furthermore, that the 4th row of A consists entirely of entries of 1 and suppose that every entry in B is 1.*

1. *What size is the matrix AB?*

2. *Write down the 4th row of AB.*

3.6 The Standard Unit Vectors and Identity Matrices

Recall that a unit vector is a vector whose magnitude is 1. A very useful set of unit vectors in R^n is the set of **standard unit vectors**, which are

$$\begin{aligned}\vec{e}_1 &= \langle 1, 0, \ldots, 0\rangle \\ \vec{e}_2 &= \langle 0, 1, \ldots, 0\rangle \\ &\vdots \\ \vec{e}_n &= \langle 0, 0, \ldots, 1\rangle .\end{aligned}$$

For each i, the vector $\vec{e}_i$ has an entry of 1 in the ith position and entries of 0 elsewhere. As a specific example, there are three standard unit vectors in R^3. They are

$$\begin{aligned}\vec{e}_1 &= \langle 1, 0, 0\rangle \\ \vec{e}_2 &= \langle 0, 1, 0\rangle \\ \vec{e}_3 &= \langle 0, 0, 1\rangle .\end{aligned}$$

It is easy to compute the dot product of any vector $\vec{x} = \langle x_1, x_2, \ldots, x_n\rangle$ in R^n with any one of the standard unit vectors in R^n. The computation is

$$\vec{x}\cdot\vec{e}_i = (x_1)(0) + (x_2)(0) + \cdots + (x_i)(1) + \cdots + (x_n)(0) = x_i.$$

Thus $\vec{x}\cdot\vec{e}_i$ is equal to the ith component of $\vec{x}$. As a specific example, if we take the vector $\vec{x} = \langle 3, 6, -2\rangle$ in R^3, then

$$\begin{aligned}\vec{x}\cdot\vec{e}_1 &= \langle 3, 6, -2\rangle \cdot \langle 1, 0, 0\rangle = 3 \\ \vec{x}\cdot\vec{e}_2 &= \langle 3, 6, -2\rangle \cdot \langle 0, 1, 0\rangle = 6 \\ \vec{x}\cdot\vec{e}_3 &= \langle 3, 6, -2\rangle \cdot \langle 0, 0, 1\rangle = -2.\end{aligned}$$

It is also easy, for any given $m \times n$ matrix A and any one of the standard unit vectors $\vec{e}_i$ in R^n, to compute the matrix–vector product $A\vec{e}_i$. The computation of $A\vec{e}_i$ using (3.7) is

$$A\vec{e}_i = (0)\operatorname{Col}_1(A) + (0)\operatorname{Col}_2(A) + \cdots + (1)\operatorname{Col}_i(A) + \cdots + (0)\operatorname{Col}_n(A) = \operatorname{Col}_i(A).$$

Likewise, if $\vec{e}_i$ is a standard unit vector in R^m, then (3.10) gives

$$A^T\vec{e}_i = (0)\operatorname{Row}_1(A) + (0)\operatorname{Row}_2(A) + \cdots + (1)\operatorname{Row}_i(A) + \cdots + (0)\operatorname{Row}_m(A) = \operatorname{Row}_i(A)$$

The facts

$$\vec{x} \cdot \vec{e}_i = x_i \tag{3.12}$$
$$A\vec{e}_i = \mathrm{Col}_i(A) \tag{3.13}$$
$$A^T\vec{e}_i = \mathrm{Row}_i(A) \tag{3.14}$$

are useful facts to remember.

The $n \times n$ **identity matrix** is the $n \times n$ matrix, denoted by I_n, whose column vectors are the standard unit vectors of R^n. Thus

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Notice that $\mathrm{Col}_i(I_n) = \vec{e}_i$ for all i and also $\mathrm{Row}_i(I_n) = \vec{e}_i$ for all i.

The most important feature of identity matrices, a feature that we will use repeatedly throughout our study of linear algebra, is that identity matrices serve as **multiplicative identity elements** for matrix multiplication. What this means is that if A is any $m \times n$ matrix, then

$$AI_n = A$$

and

$$I_m A = A.$$

Stated in words, multiplying the $m \times n$ matrix A on the right by I_n gives A and multiplying A on the left by I_m gives A.

To see why $AI_n = A$, observe that for any i $(1 \leq i \leq n)$ it follows from (3.8) and (3.13) that

$$\mathrm{Col}_i(AI_n) = A\,\mathrm{Col}_i(I_n) = A\vec{e}_i = \mathrm{Col}_i(A)$$

which shows that every column of AI_n is equal to the corresponding column of A.

To see why $I_m A = A$, observe that for any i $(1 \leq i \leq m)$ it follows from (3.11) and (3.14) that

$$\mathrm{Row}_i(I_m A) = A^T\,\mathrm{Row}_i(I_m) = A^T\vec{e}_i = \mathrm{Row}_i(A).$$

Thus every row of $I_m A$ is equal to the corresponding row of A.

Exercise 3.6.1. *1. Write down the four standard unit vectors in R^4.*

2. *Let $\{\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\}$ be the set of standard unit vectors in R^4. Let $\vec{x} = \langle 3, -3, -5, 1 \rangle$. Compute $\vec{x} \cdot \vec{e}_1$, $\vec{x} \cdot \vec{e}_2$, $\vec{x} \cdot \vec{e}_3$, and $\vec{x} \cdot \vec{e}_4$. Write out the computations in detail. You should observe that $\vec{x} \cdot \vec{e}_i = x_i$ for all $i = 1, 2, 3, 4$.*

3. *Let $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ be the set of standard unit vectors in R^3 and let A be the matrix*
$$A = \begin{bmatrix} 3 & 0 & 1 \\ 3 & 1 & -1 \\ 2 & 0 & -1 \\ 1 & 0 & -2 \end{bmatrix}.$$
Compute $A\vec{e}_1$, $A\vec{e}_2$,and $A\vec{e}_3$. Write out the computations in detail. You should observe that $A\vec{e}_i = \mathrm{Col}_i(A)$ for all $i = 1, 2, 3$.

4. *Let $\{\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\}$ be the set of standard unit vectors in R^4 and let A be the matrix given in problem 3. Compute $A^T\vec{e}_1$, $A^T\vec{e}_2$,$A^T\vec{e}_3$, and $A^T\vec{e}_4$. Write out the computations in detail. You should observe that $A^T\vec{e}_i = \mathrm{Row}_i(A)$ for all $i = 1, 2, 3, 4$.*

5. *Let A be the matrix given in problem 3 and let I_3 be the 3×3 identity matrix and let I_4 be the 4×4 identity matrix. Verify by computation that $AI_3 = A$ and $I_4A = A$.*

6. *Let $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ be the set of standard unit vectors in R^3. Then*
$$\begin{aligned} \vec{e}_1 \cdot \vec{e}_1 &= \text{\_\_\_\_} \\ \vec{e}_1 \cdot \vec{e}_2 &= \text{\_\_\_\_} \\ \vec{e}_1 \cdot \vec{e}_3 &= \text{\_\_\_\_} \\ \vec{e}_2 \cdot \vec{e}_2 &= \text{\_\_\_\_} \\ \vec{e}_2 \cdot \vec{e}_3 &= \text{\_\_\_\_} \\ \vec{e}_3 \cdot \vec{e}_3 &= \text{\_\_\_\_}. \end{aligned}$$

7. *Let $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ be the set of standard unit vectors in R^n. Then*
$$\vec{e}_i \cdot \vec{e}_i = \text{\_\_\_\_} \quad \textit{for all } i = 1, 2, \ldots, n$$
and
$$\vec{e}_i \cdot \vec{e}_j = \text{\_\_\_\_} \quad \textit{for all } i \textit{ and } j \textit{ with } i \neq j.$$

8. *Suppose that* A *is a matrix of size* 3×3, B *is a matrix of size* 3×3 *and that* $AB = I_3$. *Then*

$$A\,\mathrm{Col}_1(B) = \_\_\_\_\_\_\_$$
$$A\,\mathrm{Col}_2(B) = \_\_\_\_\_\_\_$$
$$A\,\mathrm{Col}_3(B) = \_\_\_\_\_\_\_.$$

Likewise,

$$B^T\,\mathrm{Row}_1(A) = \_\_\_\_\_\_\_\_$$
$$B^T\,\mathrm{Row}_2(A) = \_\_\_\_\_\_\_\_$$
$$B^T\,\mathrm{Row}_3(A) = \_\_\_\_\_\_\_\_.$$

9. *Suppose that* A *is a* 3×3 *matrix and suppose that* B *is a* 3×3 *matrix such that*

$$A\,\mathrm{Col}_1(B) = \vec{e}_1$$
$$A\,\mathrm{Col}_2(B) = \vec{e}_2$$
$$A\,\mathrm{Col}_3(B) = \vec{e}_3$$

where $\vec{e}_1$, $\vec{e}_2$, *and* $\vec{e}_3$ *are the standard unit vectors in* R^3.

Then $AB = \_\_\_\_\_\_\_$.

Hint: Use (3.8) to compute $\mathrm{Col}_1(AB)$, $\mathrm{Col}_2(AB)$, *and* $\mathrm{Col}_3(AB)$.

3.7 The Associative Property of Matrix Multiplication

Although matrix multiplication is not commutative, an important algebraic property of matrix multiplication that does hold true is the **associative property**. You are familiar with the associative property of multiplication of real numbers. It says that if a, b, and c are any real numbers, then $(ab)\,c = a\,(bc)$. In other words, it does not matter whether we first compute ab and then multiply the answer by c or whether we compute bc first and then multiply the answer by a. We get the same answer either way. As an example,

$$(3 \times 6) \times 2 = 18 \times 2 = 36$$

and

$$3 \times (6 \times 2) = 3 \times 12 = 36.$$

The associative property of matrix multiplication tells us that if A is an $m \times p$ matrix and B is a $p \times q$ matrix and C is $q \times n$ matrix, then

$$(AB)\,C = A\,(BC)\,.$$

Note that since A has size $m \times p$ and B has size $p \times q$, then the product AB is defined and it has size $m \times q$. Since AB has size $m \times q$ and C has size $q \times n$, then $(AB)\,C$ is also defined and it has size $m \times n$. Also, since B has size $p \times q$ and C has size $q \times n$, then BC is defined and has size $p \times n$. Since A has size $m \times p$ and BC has size $p \times n$, then $A\,(BC)$ is defined and has size $m \times n$. This reasoning tells us that $(AB)\,C$ and $A\,(BC)$ are both defined and both have the same size $(m \times n)$. Of course, the reasoning does not tell us that $(AB)\,C = A\,(BC)$. Before we look at a proof of the associative property, let us look at an example that illustrates the property.

Example 3.7.1. *Let A, B and C be the matrices*

$$A = \begin{bmatrix} -1 & -2 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -2 & 1 \\ -1 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} -2 & 1 & 0 & -2 \\ 2 & -2 & -1 & -1 \end{bmatrix}.$$

We will verify by computation that $(AB)\,C = A\,(BC)$. First we compute

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

and then compute

$$(AB)\,C = \begin{bmatrix} 4 & -5 \\ -1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -2 & 1 & 0 & -2 \\ 2 & -2 & -1 & -1 \end{bmatrix} = \begin{bmatrix} -18 & 14 & 5 & -3 \\ 6 & -5 & -2 & 0 \\ 0 & -1 & -1 & -3 \end{bmatrix}.$$

Now we compute

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

and

$$A\left(BC\right)=\left[\begin{array}{cc}-1 & -2\\ 0 & 1\\ -1 & 1\end{array}\right]\left[\begin{array}{cccc}6 & -4 & -1 & 3\\ 6 & -5 & -2 & 0\end{array}\right]=\left[\begin{array}{cccc}-18 & 14 & 5 & -3\\ 6 & -5 & -2 & 0\\ 0 & -1 & -1 & -3\end{array}\right].$$

We observe that $\left(AB\right)C=A\left(BC\right)$.

Exercise 3.7.1. *For each of the following sets of three matrices, verify by computation that* $\left(AB\right)C=A\left(BC\right)$.

1.

$$A=\left[\begin{array}{cc}3 & -1\\ -1 & -3\end{array}\right],\quad B=\left[\begin{array}{cc}2 & 3\\ -1 & -2\end{array}\right],\quad C=\left[\begin{array}{cc}1 & -3\\ 2 & 3\end{array}\right]$$

2.

$$A=\left[\begin{array}{cc}-2 & -2\\ 0 & 0\end{array}\right],\quad B=\left[\begin{array}{ccc}0 & -3 & -1\\ 2 & 1 & 0\end{array}\right],\quad C=\left[\begin{array}{ccc}1 & -2 & 3\\ -1 & -1 & -1\\ 0 & -1 & -3\end{array}\right]$$

Now let us look at a general proof of the associative property of matrix multiplication:

Suppose that A is a matrix of size $m\times p$, B is a matrix of size $p\times q$, and C is a matrix of size $q\times n$. We will prove that $\left(AB\right)C=A\left(BC\right)$ by showing that the corresponding columns of each of the matrices $\left(AB\right)C$ and $A\left(BC\right)$ are equal to each other. Each of these matrices has n columns, so we need to show that $\mathrm{Col}_i\left(\left(AB\right)C\right)=\mathrm{Col}_i\left(A\left(BC\right)\right)$ for all i between 1 and n. If we choose any i between 1 and n, then we know from (3.8) that

$$\mathrm{Col}_i\left(\left(AB\right)C\right)=\left(AB\right)\mathrm{Col}_i\left(C\right)$$

and

$$\mathrm{Col}_i\left(A\left(BC\right)\right)=A\,\mathrm{Col}_i\left(BC\right).$$

We now observe that

$$\begin{aligned}
\mathrm{Col}_i\left((AB)\,C\right) &= (AB)\,\mathrm{Col}_i\left(C\right) \\
&= c_{1i}\,\mathrm{Col}_1\left(AB\right) + c_{2i}\,\mathrm{Col}_2\left(AB\right) + \cdots + c_{qi}\,\mathrm{Col}_q\left(AB\right) \\
&= c_{1i}\left(A\,\mathrm{Col}_1\left(B\right)\right) + c_{2i}\left(A\,\mathrm{Col}_2\left(B\right)\right) + \cdots + c_{qi}\left(A\,\mathrm{Col}_q\left(B\right)\right) \\
&= A\left(c_{1i}\,\mathrm{Col}_1\left(B\right)\right) + A\left(c_{2i}\,\mathrm{Col}_2\left(B\right)\right) + \cdots + A\left(c_{qi}\,\mathrm{Col}_q\left(B\right)\right) \\
&= A\left(c_{1i}\,\mathrm{Col}_1\left(B\right) + c_{2i}\,\mathrm{Col}_2\left(B\right) + \cdots + c_{qi}\,\mathrm{Col}_q\left(B\right)\right) \\
&= A\left(B\,\mathrm{Col}_i\left(C\right)\right) \\
&= A\,\mathrm{Col}_i\left(BC\right) \\
&= \mathrm{Col}_i\left(A\left(BC\right)\right).
\end{aligned}$$

Notice that the proof of the associative law given above required us to use many different things that we have studied so far in this chapter. If you read the proof carefully, you will see that the proof required us to use all of the properties (3.8), (3.7), (3.3), and (3.4).

An important observation that we would like to make at this point is that the associative property also holds when we are dealing with a product of two matrices and a vector (rather than three matrices). This is because we have defined multiplication of a vector by a matrix in such a way that it is compatible with multiplication of two matrices. If A is a matrix of size $m \times p$, B is a matrix of size $p \times n$ and $\vec{x}$ is a vector in R^n, then

$$(AB)\,\vec{x} = A\left(B\vec{x}\right).$$

This version of the associative property will be seen to be useful in our study of linear transformations in Chapter 5.

3.8 Matrix Equations

Thus far we in this chapter, we have studied the operations of addition and subtraction of two matrices, multiplication of a matrix by a scalar, multiplication of two matrices, transposition of a matrix, and multiplication of a vector by a matrix. Another important operation that we want to study is the operation of inversion of a matrix. In order to do this, we first need to study algebraic equations that involve unknown vectors or unknown matrices.

Specifically, for a given $m \times n$ matrix A, we will study the equation

$$A\vec{x} = \vec{y} \tag{3.15}$$

where $\vec{y}$ is a vector and $\vec{x}$ is regarded as the "unknown" vector, and we will also study the equation

$$AX = Y \tag{3.16}$$

where Y is a matrix and X is regarded as the "unknown" matrix.

3.8.1 The Matrix–Vector Equation $A\vec{x} = \vec{y}$

Suppose that $A = [a_{ij}]$ is a given $m \times n$ matrix and consider the matrix–vector equation

$$A\vec{x} = \vec{y}$$

where $\vec{y} = \langle y_1, y_2, \ldots, y_m \rangle$ is a vector in R^m. Any vector $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle$ in R^n that satisfies this equation is called a **solution** of the equation and the set of all vectors $\vec{x}$ in R^n that satisfy the equation is called the **solution set** of the equation. A piece of good news is that how to solve the equation $A\vec{x} = \vec{y}$ is actually a problem we have already studied in Chapter 2! To see why, note that $A\vec{x} = \vec{y}$, when written out in detail using (3.6), is

$$\langle \text{Row}_1 (A) \cdot \vec{x}, \text{Row}_2 (A) \cdot \vec{x} \ldots, \text{Row}_m (A) \cdot \vec{x} \rangle = \langle y_1, y_2, \ldots, y_m \rangle .$$

Thus, in order for $\vec{x}$ to be a solution of $A\vec{x} = \vec{y}$, we need to have $\vec{x}$ satisfy all of the equations

$$\begin{aligned} \text{Row}_1 (A) \cdot \vec{x} &= y_1 \\ \text{Row}_2 (A) \cdot \vec{x} &= y_2 \\ &\vdots \\ \text{Row}_m (A) \cdot \vec{x} &= y_m \end{aligned}$$

which, when written out in complete detail, means that $\vec{x}$ must be solution of the system of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= y_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= y_m. \end{aligned} \tag{3.17}$$

We see that $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle$ is a solution of the matrix–vector equation $A\vec{x} = \vec{y}$ if and only if $(x_1, x_2, \ldots, x_n)$ is a solution of the system of linear equations (3.17), which has coefficient matrix $A = [a_{ij}]$. We can thus translate Theorem 2.4.1 into the following equivalent theorem for the matrix–vector equation.

Theorem 3.8.1. *Suppose that A is an $m \times n$ matrix and that $\vec{y}$ is a vector in R^m and consider the matrix–vector equation*

$$A\vec{x} = \vec{y}.$$

Let $\hat{A}$ be the augmented matrix $\hat{A} = \left[\begin{array}{c|c} A & \vec{y} \end{array}\right]$.

1. *If the rightmost column of $\hat{A}$ is a pivot column of $\hat{A}$, then $A\vec{x} = \vec{y}$ is inconsistent.*

2. *If the rightmost column of $\hat{A}$ is not a pivot column of $\hat{A}$, then $A\vec{x} = \vec{y}$ is consistent.*

Moreover, if $A\vec{x} = \vec{y}$ is consistent then

1. *If every column of A is a pivot column of A, then $A\vec{x} = \vec{y}$ has a unique solution.*

2. *If at least one column of A is not a pivot column of A, then $A\vec{x} = \vec{y}$ has infinitely many solutions.*

Example 3.8.1. *Let*

$$A = \begin{bmatrix} 3 & 0 \\ -4 & 2 \end{bmatrix} \quad \textit{and} \quad \vec{y} = \langle 6, -10 \rangle .$$

The equation $A\vec{x} = \vec{y}$ is equivalent to the system of equations

$$\begin{aligned} 3x_1 + 0x_2 &= 6 \\ -4x_1 + 2x_2 &= -10. \end{aligned} \tag{3.18}$$

To solve this system, we form the augmented matrix

$$\hat{A} = \left[\begin{array}{c|c} A & \vec{y} \end{array}\right] = \left[\begin{array}{cc|c} 3 & 0 & 6 \\ -4 & 2 & -10 \end{array}\right]$$

and perform row reduction to obtain

$$\operatorname{rref}\left(\hat{A}\right) = \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & -1 \end{array}\right].$$

We see that the system of equations (3.18) has the unique solution $(x_1, x_2) = (2, -1)$ and thus the equation $A\vec{x} = \vec{y}$ has the unique solution $\vec{x} = \langle 2, -1 \rangle$.

Example 3.8.2. *In this example, we will find the solution set of* $A\vec{x} = \vec{y}$ *where*

$$A = \begin{bmatrix} -3 & -3 \\ -2 & 3 \\ -1 & -1 \end{bmatrix} \quad \textit{and} \quad \vec{y} = \langle -3, 2, 3 \rangle .$$

The augmented matrix for this equation is

$$\hat{A} = \left[\begin{array}{c|c} A & \vec{y} \end{array}\right] = \left[\begin{array}{cc|c} -3 & -3 & -3 \\ -2 & 3 & 2 \\ -1 & -1 & 3 \end{array}\right]$$

and

$$\operatorname{rref}\left(\hat{A}\right) = \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right].$$

By looking at $\operatorname{rref}\left(\hat{A}\right)$*, we see that the rightmost column of* $\hat{A}$ *is a pivot column and thus* $A\vec{x} = \vec{y}$ *is inconsistent by Theorem 3.8.1.*

Example 3.8.3. *Let us solve* $A\vec{x} = \vec{y}$ *where*

$$A = \begin{bmatrix} 3 & 0 & 1 \\ 3 & 1 & -1 \end{bmatrix} \quad \textit{and} \quad \vec{y} = \langle 5, 7 \rangle .$$

The equation $A\vec{x} = \vec{y}$ *is equivalent to the system*

$$\begin{aligned} 3x_1 + 0x_2 + x_3 &= 5 \\ 3x_1 + x_2 - x_3 &= 7. \end{aligned}$$

The augmented matrix for this system is

$$\hat{A} = \left[\begin{array}{c|c} A & \vec{y} \end{array}\right] = \left[\begin{array}{ccc|c} 3 & 0 & 1 & 5 \\ 3 & 1 & -1 & 7 \end{array}\right]$$

and row reduction gives

$$\operatorname{rref}\left(\hat{A}\right) = \left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{3} & \frac{5}{3} \\ 0 & 1 & -2 & 2 \end{array}\right].$$

We see that x_3 is a free variable and that x_1 and x_2 are basic variables and that the solution set of the above system is given by

$$\begin{aligned} x_1 &= \frac{5}{3} - \frac{1}{3}t \\ x_2 &= 2 + 2t \\ x_3 &= t \end{aligned}$$

where t can be any real number. This means that any vector of the form

$$\vec{x} = \left\langle \frac{5}{3} - \frac{1}{3}t, 2 + 2t, t \right\rangle$$

is a solution of $A\vec{x} = \vec{y}$. We can also write the solution in the form

$$\vec{x} = \left\langle \frac{5}{3}, 2, 0 \right\rangle + t \left\langle -\frac{1}{3}, 2, 1 \right\rangle . \tag{3.19}$$

We conclude that the solution set of $A\vec{x} = \vec{y}$ consists of all vectors in R^3 that have the form (3.19) where t can be any real number. Hence there are infinitely many solutions of $A\vec{x} = \vec{y}$.

Exercise 3.8.1. *Find the solution set of $A\vec{x} = \vec{y}$ for each of the following A and $\vec{y}$.*

1.

$$A = \begin{bmatrix} 2 & 1 \\ -1 & -2 \end{bmatrix} \text{ and } \vec{y} = \langle 2, 2 \rangle$$

2.

$$A = \begin{bmatrix} 4 & 1 & 2 \\ 4 & 2 & 0 \\ 3 & 1 & 0 \end{bmatrix} \text{ and } \vec{y} = \langle -12, -12, -8 \rangle$$

3.

$$A = \begin{bmatrix} 0 & 2 & 3 \\ 3 & -1 & 2 \\ 2 & 2 & 3 \\ 3 & -2 & 3 \end{bmatrix} \text{ and } \vec{y} = \langle -1, 1, -3, 4 \rangle$$

4.

$$A = \begin{bmatrix} 0 & 2 & 3 \\ 3 & -1 & 2 \\ 2 & 2 & 3 \\ 3 & -2 & 3 \end{bmatrix} \quad \textit{and} \ \ \vec{y} = \langle 1, -1, -3, 4 \rangle$$

5.

$$A = \begin{bmatrix} -4 & 2 & -2 & -4 \\ 4 & 0 & 0 & 2 \\ 3 & -4 & 2 & -3 \end{bmatrix} \quad \textit{and} \ \ \vec{y} = \langle -6, 4, 3 \rangle$$

6.

$$A = \begin{bmatrix} -4 & -12 & -8 & -4 \\ 4 & 10 & 4 & 2 \\ 3 & 3 & -6 & -3 \end{bmatrix} \quad \textit{and} \ \ \vec{y} = \langle -6, 4, 3 \rangle$$

7.

$$A = \begin{bmatrix} 2 & 3 \\ 3 & 3 \\ 3 & 1 \\ 3 & -1 \end{bmatrix} \quad \textit{and} \ \ \vec{y} = \langle 4, 3, -1, -5 \rangle$$

8.

$$A = \begin{bmatrix} -1 & 1 \end{bmatrix} \quad \textit{and} \ \ \vec{y} = \langle 4 \rangle$$

9.

$$A = \begin{bmatrix} -3 \\ -2 \end{bmatrix} \quad \textit{and} \ \ \vec{y} = \langle 6, 4 \rangle$$

10.

$$A = \begin{bmatrix} 1 & 4 & 0 \\ 3 & -1 & 4 \\ 1 & 2 & 3 \end{bmatrix} \quad \textit{and} \ \ \vec{y} = \langle 0, 0, 0 \rangle$$

Theorem 3.8.1 tells us how to get information about the solution set of the matrix–vector equation $A\vec{x} = \vec{y}$ by using the augmented matrix $\hat{A} = \left[\begin{array}{c|c} A & \vec{y} \end{array} \right]$. In order to use the theorem, we need to be able to first write down the augmented matrix. The next theorem we will provide tells us what information can be gained if we only know the coefficient matrix A and are not given any specific vector $\vec{y}$ to serve as the right hand side of $A\vec{x} = \vec{y}$. In this case we cannot write down the augmented matrix, $\hat{A}$, and hence cannot find $\text{rref}\left(\hat{A}\right)$. However we can still write down A and find $\text{rref}(A)$. What good does that do? Well, if we just write a generic (not specific) vector $\vec{y}$ as the right hand side of $A\vec{x} = \vec{y}$, then the augmented matrix is $\hat{A} = \left[\begin{array}{c|c} A & \vec{y} \end{array}\right]$ and when the row reduction algorithm is applied to $\hat{A}$ we get

$$\text{rref}\left(\hat{A}\right) = \text{rref}\left(\left[\begin{array}{c|c} A & \vec{y} \end{array}\right]\right) = \left[\begin{array}{c|c} \text{rref}(A) & \vec{z} \end{array}\right] \tag{3.20}$$

where $\vec{z}$ is some vector that will depend on what $\vec{y}$ is. Nonetheless, since we know exactly what $\text{rref}(A)$ is, then we have complete knowledge of the pivots of A. There are four things we can conclude from knowledge of the pivots of A:

1. If every row of A contains a pivot, then it is not possible that the rightmost column of $\hat{A}$ is a pivot column of $\hat{A}$ because every row of $\text{rref}\left(\hat{A}\right)$ contains a row–leading 1 that occurs before the rightmost column of $\text{rref}\left(\hat{A}\right)$. Since $\hat{A}$ does not have a pivot in its rightmost column, then the equation $A\vec{x} = \vec{y}$ is consistent *no matter what* $\vec{y}$ is.

2. If at least one row of A does not contain a pivot, then $\text{rref}(A)$ has one or more rows of zeros at the bottom and it is possible that $\hat{A}$ will contain a pivot in its rightmost column. Whether or not that is the case depends on what $\vec{y}$ is. For some choices of $\vec{y}$, the vector $\vec{z}$ in (3.20) will be a pivot column of $\hat{A}$ and for other choices of $\vec{y}$, the vector $\vec{z}$ will not be a pivot column of $\hat{A}$. This means that for some choices of $\vec{y}$, the equation $A\vec{x} = \vec{y}$ will be inconsistent, and for other choices of $\vec{y}$ the equation $A\vec{x} = \vec{y}$ will be consistent. One particular choice of $\vec{y}$ for which consistency is guaranteed is $\vec{y} = \vec{0}_m$, because $A\vec{0}_n = \vec{0}_m$.

3. If every column of A is a pivot column of A and $A\vec{x} = \vec{y}$ is consistent, then it must be the case that $A\vec{x} = \vec{y}$ has a unique solution. This is because if every column of A is a pivot column of A, then there are no free variables.

4. If at least one column of A is not a pivot column of A and $A\vec{x} = \vec{y}$ is consistent, then $A\vec{x} = \vec{y}$ has infinitely many solutions, because free variables are present.

The above observations are summarized in the following theorem.

Theorem 3.8.2. *Suppose that A is an $m \times n$ matrix and consider the family of all matrix–vector equations of the form*

$$A\vec{x} = \vec{y}.$$

1. *If A has a pivot in every row, then $A\vec{x} = \vec{y}$ is consistent for any choice of the vector $\vec{y}$ in R^m.*
2. *If A does not have a pivot in every row, then there are some vectors $\vec{y}$ in R^m for which $A\vec{x} = \vec{y}$ is consistent and there are also some vectors $\vec{y}$ in R^m for which $A\vec{x} = \vec{y}$ is inconsistent.*

Moreover, if $\vec{y}$ is a vector such that $A\vec{x} = \vec{y}$ is consistent then

1. *If every column of A is a pivot column of A, then $A\vec{x} = \vec{y}$ has a unique solution.*
2. *If at least one column of A is not a pivot column of A, then $A\vec{x} = \vec{y}$ has infinitely many solutions.*

Example 3.8.4. *Let A be the matrix*

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

What information can we gain from A concerning solutions of the family of equations $A\vec{x} = \vec{y}$?

Theorem 3.8.2 tells us how to answer this question. We compute

$$\text{rref}\,(A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

By looking at rref (A)*, we see that every row of A contains a pivot. Theorem 3.8.2 tells us that the equation $A\vec{x} = \vec{y}$ is consistent no matter what we choose for the right hand side vector $\vec{y}$ in R^3. Furthermore, since every column of A contains a pivot, the equation $A\vec{x} = \vec{y}$ will have a unique solution (no matter what we choose as $\vec{y}$).*

Example 3.8.5. *Let A be the matrix*

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

What information can we gain from A concerning solutions of the family of equations $A\vec{x} = \vec{y}$?

To answer this question, we compute

$$\text{rref}\,(A) = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

By looking at $\text{rref}\,(A)$*, we see that every row of A contains a pivot. Theorem 3.8.2 tells us that the equation $A\vec{x} = \vec{y}$ is consistent no matter what we choose for the right hand side vector $\vec{y}$ in R^3. However, the fact that not every column of A contains a pivot (the second column of A does not contain a pivot), tells us that $A\vec{x} = \vec{y}$ will have infinitely many solutions (for any choice of $\vec{y}$ in R^3).*

Exercise 3.8.2. *For each of the following $m \times n$ matrices, A, decide whether*

a) *The equation $A\vec{x} = \vec{y}$ is consistent for any choice of $\vec{y}$ in R^m or*

b) *There are some vectors $\vec{y}$ in R^m for which $A\vec{x} = \vec{y}$ is consistent and other vectors $\vec{y}$ in R^m for which $A\vec{x} = \vec{y}$ is inconsistent*

Assuming that $\vec{y}$ is a vector in R^m such that $A\vec{x} = \vec{y}$ is consistent, decide whether

a) *$A\vec{x} = \vec{y}$ has a unique solution or*

b) *$A\vec{x} = \vec{y}$ has infinitely many solutions.*

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

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

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

4.

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

5.

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

6.

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

7.

$$A = \begin{bmatrix} 1 & 8 & -1 \\ -2 & -17 & 1 \\ 2 & 18 & 0 \end{bmatrix}$$

If A is a square matrix, then there is a simple criterion that can be applied in studying the possible solution sets of $A\vec{x} = \vec{y}$. Recall that a square matrix is a matrix that has the same number of rows as it has columns. If A is a square matrix and A has a pivot in every row, then A also has a pivot in every column (because A has the same number of rows as columns). Likewise, if A is a square matrix and at least one row of A does not contain a pivot, then at least one column of A does not contain a pivot. If A is a square matrix of size $n \times n$ and $\operatorname{rref}(A) = I_n$ (recall that I_n denotes the $n \times n$ identity matrix which was introduced in Section 3.6), then A has exactly n pivots. If $\operatorname{rref}(A) \neq I_n$, then A has fewer than n pivots. This provides us with the following corollary to Theorem 3.8.2 which applies to square coefficient matrices.

Corollary 3.8.1. *Suppose that A is an $n \times n$ (square) matrix and consider the family of all matrix–vector equations of the form*

$$A\vec{x} = \vec{y}.$$

1. *If $\operatorname{rref}(A) = I_n$, then $A\vec{x} = \vec{y}$ has a unique solution for any choice of the vector $\vec{y}$ in R^n.*

2. *If* rref $(A) \neq I_n$*, then there are some vectors* $\vec{y}$ *in* R^n *for which* $A\vec{x} = \vec{y}$ *is consistent and there are also some vectors* $\vec{y}$ *in* R^n *for which* $A\vec{x} = \vec{y}$ *is inconsistent. If* $A\vec{x} = \vec{y}$ *is consistent, then it has infinitely many solutions.*

Example 3.8.6. *Let* A *be the matrix*

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

What does Corollary 3.8.1 tell us about matrix–vector equations, $A\vec{x} = \vec{y}$*, that have* A *as their coefficient matrix?*

Since

$$\text{rref}\,(A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3,$$

Corollary 3.8.1 tells us that the equation $A\vec{x} = \vec{y}$ *will always have a unique solution, no matter what we choose for* $\vec{y}$ *on the right hand side of the equation.*

Example 3.8.7. *Let* A *be the matrix*

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

What does Corollary 3.8.1 tell us about matrix–vector equations, $A\vec{x} = \vec{y}$*, that have* A *as their coefficient matrix?*

Since

$$\text{rref}\,(A) = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & \frac{8}{3} \\ 0 & 0 & 0 \end{bmatrix} \neq I_3,$$

Corollary 3.8.1 tells us that there are some vectors $\vec{y}$ *in* R^3 *for which the equation* $A\vec{x} = \vec{y}$ *is inconsistent, and there are some vectors* $\vec{y}$ *in* R^3 *for which the equation* $A\vec{x} = \vec{y}$ *is consistent. For those* $\vec{y}$ *for which* $A\vec{x} = \vec{y}$ *is consistent,* $A\vec{x} = \vec{y}$ *has infinitely many solutions.*

Exercise 3.8.3. *For each of the following matrices* A*, what information can be gained from Corollary 3.8.1 regarding the family of equations* $A\vec{x} = \vec{y}$ *that have* A *as their coefficient matrix?*

1.

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

2.

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

3.

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

4.

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

5.

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

3.8.2 The Matrix Equation $AX = Y$

We now consider matrix equations of the form $AX = Y$ where A is an $m \times n$ matrix, X is a $n \times p$ matrix and Y is an $m \times p$ matrix. We regard A and Y as being the "given" matrices and X as being the "unknown" matrix. What we learned about matrix–vector equations of the form $A\vec{x} = \vec{y}$ in Section 3.8.1 will help us to understand the more general equations of the form $AX = Y$. If X is a matrix that is a solution of the equation $AX = Y$, then for any column of AX (say the ith column) we have

$$\mathrm{Col}_i\,(AX) = \mathrm{Col}_i\,(Y)$$

and thus by (3.8) we have

$$A\,\mathrm{Col}_i\,(X) = \mathrm{Col}_i\,(Y)\,. \tag{3.21}$$

For each i $(1 \leq i \leq n)$, equation (3.21) is a matrix–vector equation with unknown vector $\mathrm{Col}_i\,(X)$. If, for each i, we let $\hat{A}_i$ denote the augmented matrix

$$\hat{A}_i = \left[\begin{array}{c|c} A & \mathrm{Col}_i\,(Y) \end{array}\right],$$

then Theorem 3.8.1 tells us that

1. If the rightmost column of $\hat{A}_i$ is a pivot column of $\hat{A}_i$, then (3.21) is inconsistent.

2. If the rightmost column of $\hat{A}_i$ is not a pivot column of $\hat{A}_i$, then (3.21) is consistent.

Moreover, if (3.21) is consistent then

1. If every column of A is a pivot column of A, then (3.21) has a unique solution.

2. If at least one column of A is not a pivot column of A, then (3.21) has infinitely many solutions.

In order for the equation $AX = Y$ to be consistent, the equations (3.21) must be consistent for all i. Likewise, assuming that $AX = Y$ is consistent, $AX = Y$ has a unique solution if and only if all of the equations (3.21) have unique solutions. These observations can be used to obtain theorems, analogous to Theorems 3.8.1 and 3.8.2, that answer the questions of existence and uniqueness of solutions of $AX = Y$ via study of the multiply–augmented $m \times (n + p)$ matrix

$$\hat{A} = \left[\begin{array}{c|c} A & Y \end{array}\right].$$

Rather than state the most general results (the analogues of Theorems 3.8.1 and 3.8.2), we will focus on the case that A is a square matrix of size $n \times n$ and the matrix Y is also a square matrix of size $n \times n$. This is the only case that is of interest to us for the purpose of defining matrix inverses in the next section. If A and Y are both matrices of size $n \times n$, then the unknown matrix, X, in the equation $AX = Y$ must also be of size $n \times n$. The multiply–augmented $n \times (2n)$ matrix of the equation $AX = Y$ is

$$\hat{A} = \left[\begin{array}{c|c} A & Y \end{array}\right]$$

and row reduction of this matrix gives

$$\operatorname{rref}\left(\hat{A}\right) = \left[\begin{array}{c|c} \operatorname{rref}(A) & Z \end{array}\right] \tag{3.22}$$

where Z is some $n \times n$ matrix. Since A is a square matrix, then either $\operatorname{rref}(A) = I_n$ or $\operatorname{rref}(A) \neq I_n$.

If $\operatorname{rref}(A) = I_n$, then

$$\operatorname{rref}\left(\hat{A}\right) = \left[\begin{array}{c|c} I_n & X \end{array}\right]$$

where X is the unique solution of the equation $AX = Y$. If $\operatorname{rref}(A) \neq I_n$, then $AX = Y$ might be inconsistent or it might be consistent – depending on what the matrix Z in equation (3.22) turns out to be after performing row reduction on $\hat{A}$. We thus obtain the following theorem.

Theorem 3.8.3. *Suppose that A and Y are both $n \times n$ (square) matrices and consider the matrix equation*

$$AX = Y.$$

Let

$$\hat{A} = \left[\begin{array}{c|c} A & Y \end{array}\right]$$

be the $n \times (2n)$ multiply-augmented matrix that corresponds to this matrix equation.

1. *If $\operatorname{rref}(A) = I_n$, then $AX = Y$ has a unique solution. Furthermore*

$$\operatorname{rref}\left(\hat{A}\right) = \left[\begin{array}{c|c} I_n & X \end{array}\right]$$

   where X is the unique solution of $AX = Y$.

2. *If $\operatorname{rref}(A) \neq I_n$, then $AX = Y$ is either inconsistent or has infinitely many solutions.*

Example 3.8.8. *Let us solve the equation $AX = Y$ where A and Y are the 2×2 matrices*

$$A = \begin{bmatrix} -1 & -2 \\ 0 & 1 \end{bmatrix} \quad \textit{and} \quad Y = \begin{bmatrix} -1 & 1 \\ -2 & 1 \end{bmatrix}.$$

To do this, we form the multiply augmented matrix

$$\hat{A} = \left[\begin{array}{cc|cc} -1 & -2 & -1 & 1 \\ 0 & 1 & -2 & 1 \end{array}\right]$$

and perform row reduction on this matrix to obtain

$$\operatorname{rref}\left(\hat{A}\right) = \left[\begin{array}{cc|cc} 1 & 0 & 5 & -3 \\ 0 & 1 & -2 & 1 \end{array}\right] = \left[\begin{array}{c|c} I_2 & X \end{array}\right].$$

Since $\operatorname{rref}(A) = I_2$*, then*

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

is the unique solution of the equation $AX = Y$*.*

We can check that X *is a solution:*

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

Exercise 3.8.4. *Solve the equation* $AX = I_2$ *where*

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

Exercise 3.8.5. *Let* A *be the matrix*

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

Explain why the equation $AX = I_2$ *has no solution.*

We conclude this section with a corollary that is analogous to Corollary 3.8.1, but which applies to matrix equations.

Corollary 3.8.2. *Suppose that* A *is an* $n \times n$ *(square) matrix and consider the family of all matrix equations of the form*

$$AX = Y.$$

1. *If* $\operatorname{rref}(A) = I_n$*, then* $AX = Y$ *has a unique solution for any choice of* $n \times n$ *matrix* Y*.*

2. *If* $\operatorname{rref}(A) \neq I_n$*, then there are some* $n \times n$ *matrices* Y *for which* $A\vec{x} = \vec{y}$ *is consistent and there are also some* $n \times n$ *matrices* Y *for which* $AX = Y$ *is inconsistent. If* $AX = Y$ *is consistent, then it has infinitely many solutions.*

3.9 Inversion of Matrices

To motivate our definition and study of the concept of the inverse of a matrix, we will take a look at the corresponding concept as it applies in the simpler setting of real numbers. The **multiplicative identity element** for real numbers is the number 1. To say that the number 1 is the multiplicative identity element for the real numbers means that if a is any real number then

$$(a)(1) = a.$$

In words, any real numbered multiplied by 1 gives that same number. Of course, since the operation of multiplication of real numbers is commutative, it is also true that if a is any real number then

$$(1)(a) = a.$$

If a is a real number, then a real number b is said to be a **multiplicative inverse** of a if

$$ab = 1.$$

Again, because multiplication of real numbers is commutative, if b is a multiplicative inverse of a, then it will also be true that

$$ba = 1.$$

If a is any real number with $a \neq 0$, then a has a unique multiplicative inverse. For example, the multiplicative inverse of 5 is $1/5$ because

$$5\left(\frac{1}{5}\right) = 1$$

and $b = 1/5$ is the only real number that satisfies the equation $5b = 1$.

In general, the multiplicative inverse of a real number $a \neq 0$ is $1/a$, which we also sometimes write using the notation a^{-1}. So, returning to our example, we can write the multiplicative inverse of the number 5 as $1/5$ or as 5^{-1}.

What if $a = 0$? In this case, a does not have a multiplicative inverse because there does not exist any real number b such that $(0)\,b = 1$.

We are now going to define what we mean by the multiplicative inverse (if it exists) of an $n \times n$ matrix A. We want our definition of this new concept to be as similar as possible to the corresponding concept for real numbers (as discussed above). To begin, recall from Section 3.6 that the $n \times n$ identity matrix, I_n, is the multiplicative identity element for the set of all $n \times n$ matrices. This means that if A is any $n \times n$ matrix then

$$AI_n = A \quad \text{and} \quad I_n A = A.$$

Since I_n is the multiplicative identity element, then we want to define an inverse of a given $n \times n$ matrix A, to be another $n \times n$ matrix B such that $AB = I_n$ and $BA = I_n$ are both true.

Definition 3.9.1. *Suppose that A is an $n \times n$ matrix. We say that A is **invertible** if there exists an $n \times n$ matrix B such that $AB = I_n$ and $BA = I_n$.*

*If B is a matrix such that $AB = I_n$ and $BA = I_n$, then we say that B is an **inverse** of A.*

Example 3.9.1. *In this example, we illustrate the definition of the inverse of a matrix. We do not show how we came up with the matrix inverse, but you will learn how to do that soon!*

Let A be the matrix

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

Let us verify that the matrix

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

is an inverse of A. To do this we need to verify $AB = I_2$ and $BA = I_2$.

By computation, we see that

$$\begin{aligned} AB &= \begin{bmatrix} -1 & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix} \\ &\begin{bmatrix} (-1)(1) + (1)(2) & (-1)(-1) + (1)(-1) \\ (-2)(1) + (1)(2) & (-2)(-1) + (1)(-1) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= I_2 \end{aligned}$$

and likewise

$$BA = \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2.$$

This shows that B is an inverse of A.

Example 3.9.2. *Let us verify that the matrix*

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

is an inverse of itself.

By computation, we see that

$$AA = \begin{bmatrix} -1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2,$$

which shows that A is an inverse of itself.

Exercise 3.9.1. *1. Verify that the matrix*

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

is an inverse of the matrix

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

2. Verify that the matrix

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

is an inverse of the matrix

$$A = \begin{bmatrix} \frac{1}{2} & -1 & -6 \\ 0 & 0 & 1 \\ 0 & \frac{1}{2} & 2 \end{bmatrix}.$$

Exercise 3.9.2. *Suppose that A is an invertible $n \times n$ matrix and suppose that B is an inverse of A. Must it also be true that A is an inverse of B? Explain.*

Our first theorem concerning matrix inverses states that an invertible matrix has exactly one inverse.

Theorem 3.9.1. *If A is an invertible $n \times n$ matrix, then the inverse of A is unique. This means that there do not exist two different $n \times n$ matrices, B and C, that both act as inverses of A.*

Proof. Suppose that A is invertible and suppose that the matrices B and C both act as inverses of A. This means that

$$AB = I_n \quad \text{and} \quad BA = I_n$$

and

$$AC = I_n \quad \text{and} \quad CA = I_n.$$

This implies that

$$B = BI_n = B\,(AC) = (BA)\,C = I_nC = C.$$

We have shown that if matrices B and C both act as inverses of A, then B and C must be identical. □

Since an invertible matrix, A, has only one inverse, we can give this inverse a name. The name that we give to it is A^{-1}. This is in keeping with the fact that we give the name a^{-1} to the inverse of an invertible real number a. In addition, we can stop using the word "an" and start using the word "the" when referring to the inverse of a matrix. In speaking and writing, we use the indefinite articles "a" and "an" when we are in situations where we think that there possibly could be more that one of something. For example, perhaps we are in the midst of some problem where we are discussing solutions of some equation that has the form $f\,(x) = b$ and we aren't sure how many solutions this equation has. In that case, we might make a statement of the form "Suppose that x is **a solution** of the equation $f\,(x) = b$." However, if we are sure that the equation we are discussing has exactly one solution, then we use the definite article "the" to refer to this solution. We would say "Suppose that x is **the solution** of the equation $f\,(x) = b$." To illustrate the A^{-1} notation, we showed in Example 3.9.1 that **the inverse** of the matrix

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

is the matrix

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

The big questions that we now focus on are

1. Which matrices are invertible?

2. How do we find the inverse of an invertible matrix?

The answers to both questions are given in the Theorem 3.9.3. Before providing that theorem, we provide a prelimnary theorem that tells us that it is actually not necessary to check that both $AB = I_n$ and $BA = I_n$ in order to conclude that $B = A^{-1}$. This is surprising! It is surprising because we know that matrix multiplication is not commutative, meaning that it is usually not true that $AB = BA$ even when AB and BA are both defined and have the same size. What the upcomig theorem tells us, though, is that if $AB = I_n$, then it must also be true that $BA = I_n$ (and incidentally $AB = BA$).

Theorem 3.9.2. *Suppose that A and B are both $n \times n$ matrices and suppose that $AB = I_n$. Then $BA = I_n$.*

Proof. Suppose that $AB = I_n$.

We know that

$$I_n Y = Y \text{ for all } n \times n \text{ matrices } Y$$

and thus

$$(AB)\,Y = Y \text{ for all } n \times n \text{ matrices } Y.$$

By the associative property of matrix multiplication, we can write the above statement as

$$A\,(BY) = Y \text{ for all } n \times n \text{ matrices } Y.$$

This shows that the equation $AX = Y$ has a solution for any choice of $n \times n$ matrix Y. (The above equation shows that BY is a solution of $AX = Y$.)

Since $AX = Y$ has a solution for any choice of Y, then by Corollary 3.8.2 it must be the case that $\operatorname{rref}(A) = I_n$ and thus it must in fact be true that $AX = Y$ has a *unique* solution for *any choice* of $n \times n$ matrix Y. Thus it must be the case that the equation $AX = A$ has unique solution.

It is easy to see that $X = I_n$ is a solution of the equation $AX = A$.

However, note that

$$A\,(BA) = (AB)\,A = I_n A = A$$

which shows that BA is *also* a solution of $AX = A$.

We have determined that the equation $AX = A$ has a unique solution, but we have also determined that $X = I_n$ and $X = BA$ both satisfy this equation. Therefore it must be the case that $BA = I_n$.

This completes the proof.

□

We now provide the theorem that answers the questions on which matrices are invertible and on how to find the inverses of invertible matrices.

Theorem 3.9.3. *Suppose that A is an $n \times n$ matrix. Then A is invertible if and only if* $\mathrm{rref}\,(A) = I_n$.

Furthermore, if $\mathrm{rref}\,(A) = I_n$ *and $\hat{A}$ is the multiply–augmented matrix*

$$\hat{A} = \left[\begin{array}{c|c} A & I_n \end{array}\right],$$

then

$$\mathrm{rref}\left(\hat{A}\right) = \left[\begin{array}{c|c} I_n & A^{-1} \end{array}\right].$$

Proof. Suppose that A is an $n \times n$ matrix and suppose that $\mathrm{rref}\,(A) = I_n$. Then consider the matrix equation $AX = I_n$. The multiply–augmented matrix for this matrix equation is

$$\hat{A} = \left[\begin{array}{c|c} A & I_n \end{array}\right]$$

and since $\mathrm{rref}\,(A) = I_n$, then statement 1 of Theorem 3.8.3 tells us that $AX = I_n$ has a unique solution and that

$$\mathrm{rref}\left(\hat{A}\right) = \left[\begin{array}{c|c} I_n & X \end{array}\right]$$

where X is the unique solution of $AX = I_n$.

We have shown that there is a unique matrix X, such that $AX = I_n$. By Theorem 3.9.2, it must also be true that $XA = I_n$. Thus A is invertible and $X = A^{-1}$. Furthermore, we see that

$$\mathrm{rref}\left(\hat{A}\right) = \left[\begin{array}{c|c} I_n & A^{-1} \end{array}\right].$$

We have proved that if $\mathrm{rref}\,(A) = I_n$, then A is invertible. We still need to prove that if $\mathrm{rref}\,(A) \neq I_n$, then A is not invertible. We will do this by proving the (equivalent) contrapositive statement: If A is invertible, then $\text{rref}(A) = I_n$.

Suppose that A is invertible. Then we know that A^{-1} exists. Knowing that A^{-1} exists allows us to easily solve the matrix equation $AX = I_n$. If $AX = I_n$, then

$$A^{-1}(AX) = A^{-1}I_n$$

which implies that

$$\left(A^{-1}A\right)X = A^{-1}$$

which implies that

$$I_nX = A^{-1}$$

which implies that

$$X = A^{-1}.$$

The fact that the matrix equation $AX = I_n$ has a unique solution (which is $X = A^{-1}$) tells us, by Theorem 3.8.3, that $\text{rref}(A) = I_n$. (If it were the case that $\text{rref}(A) \neq I_n$, then it would have to be the case that the matrix equation $AX = I_n$ either has no solution or infinitely many solutions.) □

The fact that if A is invertible, then

$$\text{rref}\left(\hat{A}\right) = \left[\begin{array}{c|c} I_n & A^{-1} \end{array}\right],$$

provides us with a procedure for calculating A^{-1} for any square matrix A. The procedure is to form the multiply–augmented matrix $\hat{A}$ and then perform the row reduction algortihm on it. The left half of $\text{rref}\left(\hat{A}\right)$ always turns out to be I_n and the right half of $\text{rref}\left(\hat{A}\right)$ always turns out to be A^{-1}. This is illustrated in the following example, followed by some exercises that are provided for practice.

Example 3.9.3. *In Example 3.9.1, we showed that the inverse of the matrix*

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

is the matrix

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

but we did not show how to find A^{-1} from A. Here is how. We form the multiply–augmented matrix

$$\hat{A} = \left[\begin{array}{c|c} A & I_2 \end{array}\right] = \left[\begin{array}{cc|cc} -1 & 1 & 1 & 0 \\ -2 & 1 & 0 & 1 \end{array}\right]$$

and then perform the row reduction algortihm on $\hat{A}$ to obtain

$$\operatorname{rref}\left(\hat{A}\right) = \left[\begin{array}{cc|cc} 1 & 0 & 1 & -1 \\ 0 & 1 & 2 & -1 \end{array}\right] = \left[\begin{array}{c|c} I_2 & A^{-1} \end{array}\right].$$

Thus we have computed

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

Exercise 3.9.3. *In Example 3.9.2, we showed that the inverse of the matrix*

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

is A. Verify that

$$\operatorname{rref}\left(\left[\begin{array}{c|c} A & I_2 \end{array}\right]\right) = \left[\begin{array}{c|c} I_2 & A \end{array}\right].$$

Exercise 3.9.4. *Use row reduction to find the inverses of the following matrices. Once you have done this, check by computation to make sure that you have gotten the right answer by computing AA^{-1} (using the A^{-1} you have found).*

1.

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

2.

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

3.

$$A = \begin{bmatrix} \frac{1}{2} & -1 & -6 \\ 0 & 0 & 1 \\ 0 & \frac{1}{2} & 2 \end{bmatrix}$$

4.

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

5.

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

The following corollary follows from Theorem 3.9.3 and the proof of the corollary is contained within the proof of the theorem.

Corollary 3.9.1. *Suppose that A is an invertible $n \times n$ matrix.*

Then, for any vector $\vec{y}$ in R^n, the matrix–vector equation $A\vec{x} = \vec{y}$ has the unique solution

$$\vec{x} = A^{-1}\vec{y}$$

and, for any $n \times p$ matrix Y, the matrix equation $AX = Y$ has the unique solution

$$X = A^{-1}Y.$$

Solving equations of the form $A\vec{x} = \vec{y}$ and $AX = Y$ by first computing A^{-1} and then applying Corollary 3.9.1 is not the most efficient way to solve these equations. The most efficient way is to use augmented matrices and row reduction directly on the equations. Nonetheless, the corollary is valuable as a theoretical tool that will be used in future parts of this course. Here is an example that illustrates the corollary.

Example 3.9.4. *Let us use matrix inversion to solve the equation $A\vec{x} = \vec{y}$ where*

$$A = \begin{bmatrix} -1 & 1 \\ -2 & 1 \end{bmatrix} \text{ and } \vec{y} = \langle 4, 3 \rangle .$$

We have already seen (in Example 3.9.1) that

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

By Corollary 3.9.1, the unique solution of $A\vec{x} = \vec{y}$ is $\vec{x} = A^{-1}\vec{y}$. Since

$$\begin{aligned} \mathrm{Row}_1\left(A^{-1}\right) \cdot \vec{y} &= \langle 1, -1\rangle \cdot \langle 4, 3\rangle = 1 \\ \mathrm{Row}_2\left(A^{-1}\right) \cdot \vec{y} &= \langle 2, -1\rangle \cdot \langle 4, 3\rangle = 5, \end{aligned}$$

then the unique solution of $A\vec{x} = \vec{y}$ is $\vec{x} = \langle 1, 5\rangle$.

Exercise 3.9.5. *For the matrices A and vectors $\vec{y}$ given below. Solve the equation $A\vec{x} = \vec{y}$ in two different ways:*
a) *by performing row reduction on the augmented matrix $\hat{A} = \left[\begin{array}{c|c} A & \vec{y} \end{array}\right]$*
b) *by first computing A^{-1} (using the algorithm for computing A^{-1}) and then using Corollary 3.9.1.*

1.
$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \vec{y} = \langle -1, -3\rangle$$

2.
$$A = \begin{bmatrix} -1 & 2 \\ -2 & 1 \end{bmatrix} \quad \text{and} \quad \vec{y} = \langle 5, 1\rangle$$

3.
$$A = \begin{bmatrix} -2 & -2 & 0 \\ 0 & 0 & -3 \\ -1 & 2 & 1 \end{bmatrix} \quad \text{and} \quad \vec{y} = \langle 4, 3, -8\rangle$$

4.
$$A = \begin{bmatrix} 3 & 1 & 0 & -1 \\ 2 & -2 & 2 & 1 \\ 1 & -1 & 3 & 3 \\ -2 & -1 & -1 & 1 \end{bmatrix} \quad \text{and} \quad \vec{y} = \langle -5, 3, 5, 7\rangle .$$

We conclude this section with two useful theorems – one stating that the product of two invertible matrices is invertible and the other stating that the transpose of an invertible matrix is invertible.

Theorem 3.9.4. *Suppose that A and B are $n \times n$ matrices and that A and B are both invertible. Then the product AB is invertible and*

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Proof. Suppose that A and B are $n \times n$ matrices and that A and B are both invertible. This means that A^{-1} and B^{-1} both exist. Now notice (by using the associative property of matrix multiplication) that

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

We have shown that $(AB)\left(B^{-1}A^{-1}\right) = I_n$ and we thus conclude that $(AB)^{-1} = B^{-1}A^{-1}$. □

Theorem 3.9.4 tells us that the inverse of a product of matrices is the product of the inverses in the reverse order. Remember that matrix multiplication is not commutative, so keeping the order as stated in the theorem is essential. It is *not* generally true that $(AB)^{-1} = A^{-1}B^{-1}$.

Theorem 3.9.5. *Suppose that A is an invertible $n \times n$ matrix. Then A^T is also invertible and $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.*

Proof. Since A is invertible, then A^{-1} exists and $A^{-1}A = I_n$. Taking the transpose of both sides of this equation gives $\left(A^{-1}A\right)^T = I_n^T$. However, $I_n^T = I_n$, and by property (3.5), we have $\left(A^{-1}A\right)^T = A^T\left(A^{-1}\right)^T$. Thus

$$A^T\left(A^{-1}\right)^T = I_n$$

and we conclude from this that A^T is invertible and that $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$. □

Example 3.9.5. *Let us illustrate Theorem 3.9.4 using the matrices*

$$A = \begin{bmatrix} 3 & 0 \\ 1 & 3 \end{bmatrix} \quad \textit{and} \quad B = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}.$$

The inverses of these matrices are

$$A^{-1} = \begin{bmatrix} \frac{1}{3} & 0 \\ -\frac{1}{9} & \frac{1}{3} \end{bmatrix} \quad \textit{and} \quad B^{-1} = \begin{bmatrix} 0 & \frac{1}{2} \\ -1 & \frac{1}{2} \end{bmatrix}.$$

The product AB is

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

and the inverse of the product is

$$(AB)^{-1} = \begin{bmatrix} -\frac{1}{18} & \frac{1}{6} \\ -\frac{7}{18} & \frac{1}{6} \end{bmatrix}.$$

We also observe that

$$B^{-1}A^{-1} = \begin{bmatrix} 0 & \frac{1}{2} \\ -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 0 \\ -\frac{1}{9} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} -\frac{1}{18} & \frac{1}{6} \\ -\frac{7}{18} & \frac{1}{6} \end{bmatrix},$$

illustrating that $(AB)^{-1} = B^{-1}A^{-1}$.

Exercise 3.9.6. *For the matrices*

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

verify by computation that $(AB)^{-1} = B^{-1}A^{-1}$.

Exercise 3.9.7. *Suppose that A and B, are $n \times n$ matrices and suppose that A is invertible and B is not invertible. Explain why AB cannot be invertible.*

Exercise 3.9.8. *Theorem 3.9.4 tells us that if A is invertible and B is invertible, then AB is invertible. Exercise 3.9.7 tells us that if A is invertible and B is not invertible, then AB is not invertible. We can ask whether there is any relationship between the invertibility of a pair of matrices A and B and their sum $A + B$. Generally speaking, there is no clear relationship. Create an example of a pair of 2×2 matrices A and B such that*

1. *A and B are invertible, and $A + B$ is invertible,*

2. *A and B are invertible, but $A + B$ is not invertible,*

3. *A and B are not invertible, but $A + B$ is invertible,*

4. *A and B are not invertible, and $A + B$ is also not invertible,*

Exercise 3.9.9. *For the square matrices, A, given below, find A^{-1} and $(A^{-1})^T$. Then verify that $\left(A^T\right)^{-1} = (A^{-1})^T$ by computing $A^T (A^{-1})^T$.*

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

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

3.10 Additional Exercises

(Jump to Solutions)

1. Complete the following sentences by filling in one of the words “scalar”, “vector”, or “matrix”.

   (a) The sum of two vectors is a __________.

   (b) The sum of two matrices is a __________.

   (c) A scalar multiple of a vector is a __________.

   (d) A scalar multiple of a matrix is a __________.

   (e) The product of two matrices is a __________.

   (f) The dot product of two vectors is a __________.

   (g) A linear combination of vectors is a __________.

   (h) The product of a matrix and a vector is a __________.

   (i) The transpose of a matrix is a __________.

   (j) The inverse of a matrix is __________.

2. Suppose that the matrix A has column vectors

$$\begin{aligned} \mathrm{Col}_1(A) &= \langle 1, -2, 7, -4 \rangle \\ \mathrm{Col}_2(A) &= \langle -1, 7, 3, -1 \rangle \\ \mathrm{Col}_3(A) &= \langle -6, 8, 6, 0 \rangle . \end{aligned}$$

   Write down A and write down the row vectors of A.

3. Suppose that the matrix B has row vectors

$$\begin{aligned}\text{Row}_1(B) &= \langle -6, 4, 4, -2 \rangle \\ \text{Row}_2(B) &= \langle 5, -2, 5, 5 \rangle \\ \text{Row}_3(B) &= \langle 3, 6, 2, -5 \rangle .\end{aligned}$$

   Write down B and write down the column vectors of B.

4. For the matrices A and B that you wrote down in questions 1 and 2 above, is it possible to compute $A + B$? If so, then compute it. If not, then explain why not? Is it possible to compute AB? If so, then compute it. If not, then explain why not.

5. The $m \times n$ zero matrix is the $m \times n$ matrix that has all entries of 0. This matrix is denoted by $O_{m\times n}$. Thus, for example,

$$O_{3\times 4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

   Explain why it makes sense to refer to $O_{m\times n}$ as the **additive identity element** for the set of all $m \times n$ matrices.

6. For the matrices

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

   compute $2(A + B)$ and $2A + 2B$ and observe that they are the same.

7. For the matrices

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

   compute AB and BA.

8. We know that it is not generally true that $AB = BA$, even when A and B are square matrices of the same size. Let A be the 2×2 matrix

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

   Find **all** matrices X such that $AX = XA$ is true.

9. Suppose that A is an $n \times n$ matrix and let $O_{n\times n}$ be the $n \times n$ zero matrix. (Refer to problem 5 above.) Explain why $AO_{n\times n} = O_{n\times n}$.

10. A property of real numbers that you are probably familiar with is that if a and b are real numbers and $ab = 0$, then it must be true that either $a = 0$ or $b = 0$. A similar property *does not* hold for matrices. Come up with an example of a 2×2 matrix A and a 2×2 matrix B such that AB is equal to the 2×2 zero matrix (that is $AB = O_{2\times 2}$) but neither A nor B is equal to $O_{2\times 2}$.

11. Another property of real numbers that you are probably familiar with is the "cancellation law" which says that if a, b, and c are real numbers with $a \neq 0$ and $ab = ac$, then $b = c$. A similar property *does not,* in general, hold for matrices. Come up with an example of 2×2 matrices A, B, and C such that $A \neq O_{2\times 2}$ and $AB = AC$ but $B \neq C$.

    Hint: If you successfully did problem 10, then you should be able to use what you got there to help with this problem.

12. Referring back to the previous problem, there **is** a cancellation law that holds when the matrix A is invertible. Specifically, if A is an invertible $n \times n$ matrix and B and C are matrices of size $n \times p$ and $AB = AC$, then $B = C$. Prove this cancellation law.

    Hint: Multiply both sides of $AB = AC$ on the left by A^{-1} and use the associative property of matrix multiplication and the fact that I_n is a multiplicative identity element for matrix multiplication.

13. Suppose that A is an $m \times n$ matrix. Explain why the matrix product AA^T is defined (is possible to carry out). What size is AA^T?

14. For the matrix
$$A = \begin{bmatrix} 3 & 0 & 5 \\ 3 & 1 & 0 \end{bmatrix},$$
    compute AA^T and $\left(AA^T\right)^T$.

    What do you observe? It is not a coincidence.

    Use property (3.5) to prove that if A is any matrix, then $\left(AA^T\right)^T = AA^T$.

15. For the matrix and vector

$$A = \begin{bmatrix} -2 & -1 & 1 & 1 \\ -2 & 2 & -2 & -1 \\ -2 & 2 & 2 & -2 \end{bmatrix} \quad \text{and} \quad \vec{x} = \langle x_1, x_2, x_3, x_4 \rangle \,,$$

Compute $A\vec{x}$ in two different ways: **a)** by using (3.6) and **b)** by using (3.7).

16. For the matrix and vector

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 2 \\ 0 & -1 & 2 \end{bmatrix} \quad \text{and} \quad \vec{x} = \langle x_1, x_2, x_3 \rangle \,,$$

Compute $A^T\vec{x}$ in four different ways: **a)** by first computing A^T and then using (3.6), **b)** by first computing A^T and then using (3.7), **c)** by using (3.9), and **d)** by using (3.10).

17. Suppose that A is an $n \times n$ matrix and suppose that the vector $\vec{x}$ in R^n is a solution of the homogeneous equation $A\vec{x} = \vec{0}_n$. Explain why all of the row vectors of A are orthogonal to $\vec{x}$.

18. Suppose that A is a 3×3 matrix and that

$$\text{rref}\left(\left[\begin{array}{c|c} A & \begin{array}{c} 2 \\ -1 \\ -2 \end{array} \end{array} \right] \right) = \left[\begin{array}{c|c} I_3 & \begin{array}{c} -1 \\ -3 \\ 2 \end{array} \end{array} \right].$$

What is the solution of the matrix–vector equation

$$A\vec{x} = \langle 2, -1, -2 \rangle \,?$$

19. For the following matrices A and vectors $\vec{y}$, find the solution set of the equation $A\vec{x} = \vec{y}$. Indicate whether the equation is inconsistent, consistent with a unique solution, or consistent with infinitely many solutions.

(a)

$$A = \begin{bmatrix} 0 & -1 & 0 \\ -1 & -1 & 0 \\ 1 & -1 & -1 \end{bmatrix} \quad \text{and} \quad \vec{y} = \langle 1, -3, 2 \rangle$$

(b)

$$A = \begin{bmatrix} 0 & -1 & 0 \\ 0 & -1 & 0 \\ 1 & -1 & -1 \end{bmatrix} \quad \text{and} \quad \vec{y} = \langle -2, -2, -2 \rangle$$

(c)

$$A = \begin{bmatrix} 0 & -1 & 0 \\ 0 & -1 & 0 \\ 1 & -1 & -1 \end{bmatrix} \quad \text{and} \quad \vec{y} = \langle 0, 1, -2 \rangle$$

20. Write down the multiply–augmented matrix for the matrix equation $AX = Y$ where A and Y are the matrices

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

and then perform row reduction on this multiply–augmented matrix to find the solution of the equation $AX = Y$. Does $AX = Y$ have a unique solution?

21. Write down the multiply–augmented matrix for the matrix equation $AX = I_2$ where A is the matrix

$$A = \begin{bmatrix} 1 & 4 \\ 3 & 0 \end{bmatrix}$$

and then perform row reduction on this multiply–augmented matrix to find the solution of the equation $AX = I_2$. (What you are doing here is finding A^{-1}.)

22. Let A be the 3×3 matrix

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

Find A^{-1} by solving $AX = I_3$.

23. Show that the matrix

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

is not invertible by studying the equation $AX = I_3$.

24. Let $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ be the set of standard unit vectors in R^3 and let $\vec{x} = \langle x_1, x_2, x_3 \rangle$ be a vector in R^3. Show that

$$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3.$$

This shows that any vector in R^3 is a linear combination of the standard unit vectors of R^3.

25. Suppose that A is an $n \times n$ matrix. It is obvious that $A\vec{0}_3 = \vec{0}_3$.

Suppose that there exists some vector $\vec{x}$ in R^3 with $\vec{x} \neq \vec{0}_3$ such that $A\vec{x} = \vec{0}_3$. Explain why it must be the case that $\text{rref}\,(A) \neq I_3$ (and hence that A is not invertible).

26. (a) Explain why an $n \times n$ matrix, A, that contains a row consisting entirely of entries of 0 is not invertible.

    Hint: Think about solving $AX = I_n$. What can you say about $\text{rref}\,(A)$?

    (b) Explain why an $n \times n$ matrix, A, that has two identical rows is not invertible.

Chapter 4

Vector Spaces and Subspaces

In this chapter, we will investigate vector spaces from the perspective of an algebraic setting. We will primarily focus on the vector spaces R^n, but we will see that we can generalize the idea of "vector" and arrive at a framework that extends to other sets of objects that we can manipulate with specific operations. We've already seen these ideas extended to matrices in Chapter 3.

We used the phrase *vector space* extensively in Chapter 1 in reference to R^n, and it is tempting to attach a physical connotation, especially to the word *space*. But when mathematicians use the phrase vector space, it refers to an algebraic structure—select objects, operations on those objects, and properties that are either assumed (i.e., axioms) or can be deduced from the assumptions (e.g., theorems).

4.1 Linear Independence

In Chapter 2, we studied systems of linear equations where we found that there are three possible solution scenarios: no solution, a unique solution, infinitely many solutions. Later in Chapter 3, we found that a system of linear equations can be formulated as a matrix–vector equation $A\vec{x} = \vec{y}$. And in section 3.5, we learned that the matrix-vector product $A\vec{x}$ can be interpreted as a linear combination of the columns of A with the entries of $\vec{x}$ as the weights, i.e.,

$$A\vec{x} = x_1 \operatorname{Col}_1(A) + x_2 \operatorname{Col}_2(A) + \cdots + x_n \operatorname{Col}_n(A). \tag{4.1}$$

Hence, for a given vector $\vec{y}$, the question

> "Does $A\vec{x} = \vec{y}$ have a solution?"

can be rephrased as the question

> "Is $\vec{y}$ in the subset of R^m spanned by the column vectors of the matrix A—i.e., is $\vec{y} \in \text{Span}\{\text{Col}_1(A), \text{Col}_2(A), \cdots, \text{Col}_n(A)\}$?"

Recalling that the homogeneous system $A\vec{x} = \vec{0}_m$ is always consistent (since it necessarily permits the trivial solution $\vec{x} = \vec{0}_n$), the critical question for a homogeneous system is whether it has nontrivial solutions. In light of equation (4.1), we can consider the existence (or not) of nontrivial solutions to be some property of the set of column vectors of A.

Of course, given any set of vectors, say $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ in R^m, we can always create a matrix A by setting, $\text{Col}_i(A) = \vec{v}_i$. So we don't have to think of this property in terms of matrices, but rather as a possible relationship among a set of vectors. This property is called **linear independence**.

Definition 4.1.1. *The collection of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ in R^m is said to be* **linearly independent** *if the homogeneous equation*

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{0}_m \tag{4.2}$$

has only the trivial solution, $x_1 = x_2 = \cdots = x_n = 0$.

If the collection of vectors is not linearly independent, then we say that it is **linearly dependent**. *For a linearly dependent set of vectors, an equation of the form (4.2) having at least one nonzero weight is called a* **linear dependence relation**.

Remark 4.1.1. *Note that no matter what the vectors are, the equation (4.2) can always be made true by simply taking all of the coefficients to be zero. The question is whether it's possible to make that equation true without insisting that all coefficients are zero. This definition can be restated to focus on linear dependence. In this case, we'll say*

> *The collection of vectors is* **linearly dependent** *if there exists a set of weights, $x_1, \ldots, x_n$,* **not all zero** *such that equation (4.2) holds.*

We can easily characterize the linear dependence or independence of a set consisting of one or two vectors. In particular, a set containing a single vector $\vec{v}$ in R^m is linearly dependent if and only if $\vec{v} = \vec{0}_m$. This is readily confirmed by considering the equation

$$c\vec{v} = \vec{0}_m. \tag{4.3}$$

If $\vec{v} \neq \vec{0}_m$, then (4.3) requires the coefficient $c = 0$, whereas if $\vec{v} = \vec{0}_m$, we can take $c = 1$ (or any other nonzero number) to produce a linear dependence relation.

For a set consisting of two vectors, $\vec{v}_1$ and $\vec{v}_2$ in R^m, the set is linearly dependent if and only if one of these is a scalar multiple of the other. To establish this, suppose $\vec{v}_1 = c\vec{v}_2$ where c is any scalar (possibly even zero). We can rearrange this to obtain the linear dependence relation

$$1\vec{v}_1 - c\vec{v}_2 = \vec{0}_m.$$

Note that at least one of these coefficients (the 1 scaling $\vec{v}_1$) is nonzero—making this a valid linear dependence relation and showing that the vectors are linearly dependent. Alternatively, if the pair is linearly dependent, then there exists scalars c_1 and c_2, not both zero such that

$$c_1\vec{v}_1 + c_2\vec{v}_2 = \vec{0}_m.$$

We can assume without loss of generality[1] that $c_1 \neq 0$ and rearrange the equation to find that $\vec{v}_1$ is a scalar multiple of $\vec{v}_2$, i.e.,

$$\vec{v}_1 = -\frac{c_2}{c_1}\vec{v}_2.$$

Exercise 4.1.1. *For each set of one or two vectors, determine whether the set is linearly dependent or linearly independent.*

1. $\vec{v} = \langle 1, 0, 1, 2 \rangle$
2. $\vec{v}_1 = \langle 1, -1 \rangle$, $\vec{v}_2 = \langle -4, 4 \rangle$
3. $\vec{v}_1 = \langle 0, 0, 0, 0, 0 \rangle$, $\vec{v}_2 = \langle 1, 2, 3, 4, 5 \rangle$

[1] To say that we can assume something without loss of generality means that we can make this assumption (e.g., $c_1 \neq 0$) without damaging the integrity of our argument. You might wonder, what if $c_1 = 0$? Since we are assuming that at least one of these numbers is nonzero, we could just swap their labels ($c_1 \leftrightarrow c_2$ and $\vec{v}_1 \leftrightarrow \vec{v}_2$) because their labels are completely arbitrary.

4. $\vec{v}_1 = \langle 3, 6, 18 \rangle$, $\vec{v}_2 = \left\langle \frac{1}{3}, \frac{2}{3}, 2 \right\rangle$

5. $\vec{v}_1 = \langle 2, 1, 0 \rangle$, $\vec{v}_2 = \langle -1, 2, 0 \rangle$

For a collection of three or more vectors, determining linear dependence will be slightly more complicated. However, equation (4.2) gives rise to a homogeneous system of linear equations, and we can use familiar techniques to investigate the existence of nontrivial solutions.

Example 4.1.1. *Determine whether the vectors* $\vec{v}_1 = \langle 1, 1, -1 \rangle$, $\vec{v}_2 = \langle 2, 0, -3 \rangle$, *and* $\vec{v}_3 = \langle 0, 2, 1 \rangle$ *are linearly independent or linearly dependent. If they are linearly dependent, construct a linear dependence relation.*

We can consider the equation $x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 = \vec{0}_3$,

$$x_1\langle 1, 1, -1 \rangle + x_2\langle 2, 0, -3 \rangle + x_3\langle 0, 2, 1 \rangle = \langle 0, 0, 0 \rangle.$$

Using the results of Chapter 3, we can restate the equation in the form of a homogeneous matrix-vector equation $A\vec{x} = \vec{0}_3$ *where* $\mathrm{Col}_i(A) = \vec{v}_i$, $i = 1, 2, 3$. *We can reduce the augmented matrix,* $[A \mid \vec{0}_3]$ *to an rref.*

$$\left[\begin{array}{rrr|r} 1 & 2 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ -1 & -3 & 1 & 0 \end{array}\right] \xrightarrow{rref} \left[\begin{array}{rrr|r} 1 & 0 & 2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

From the rref, we see that there are two basic variables, x_1 *and* x_2, *and one free variable,* x_3. *Because there is a free variable, the homogeneous system* $A\vec{x} = \vec{0}_3$ *has nontrivial solutions. Hence, we conclude that the vectors are linearly dependent.*

We can actually glean more information from the rref. In particular, the values that appear in the non-pivot column can be used to construct a linear dependence relation. From the rref, we see that $x_1 = -2x_3$, $x_2 = x_3$, *and* x_3 *is a free variable. The solution is given in parametric form as*

$$\begin{array}{rcl} x_1 & = & -2t, \\ x_2 & = & t, \qquad t \in R. \\ x_3 & = & t \end{array}$$

To construct a linear dependence relation, we can choose any nonzero value[2]

[2] Different values of t will give different linear dependence relations that are equally valid. Perhaps it would be desirable to select $t = -1$ instead to avoid the leading minus sign, $2\vec{v}_1 - \vec{v}_2 - \vec{v}_3 = \vec{0}_3$.

for the parameter t. For example, if we set $t = 1$, we obtain a linear dependence relation

$$-2\vec{v}_1 + \vec{v}_2 + \vec{v}_3 = \vec{0}_3.$$

Example 4.1.1 illustrates a general result that follows directly from the definition of linear independence.

Theorem 4.1.1. *Let A be an $m \times n$ matrix. The column vectors of A are linearly independent in R^m if and only if the homogeneous equation $A\vec{x} = \vec{0}_m$ has only the trivial solution.*

In the case of square matrices, Theorem 4.1.1 connects the linear independence of the columns to the property of invertibility.

Corollary 4.1.1. *If A is an $n \times n$ matrix, then A is invertible if and only if the columns of A are linearly independent.*

Recall Theorem 3.9.3, that invertibility is equivalent to being row equivalent to I_n which is in turn equivalent to $A\vec{x} = \vec{0}_n$ having only the trivial solution. Now, we can say that invertibility is equivalent to having linearly independent columns. Since the invertibility of A implies the invertibility of A^T (Theorem 3.9.5), we can see that if A is an invertible square matrix, then its rows are also linearly independent.

Example 4.1.2. *Determine whether the columns of $A = \begin{bmatrix} 1 & 3 & -1 \\ 0 & 4 & 2 \\ 5 & 2 & 3 \end{bmatrix}$ are linearly dependent or linearly independent. If dependent, write a linear dependence relation.*

We can consider the homogeneous equation $A\vec{x} = \vec{0}_3$. Using the augemented matrix $[A\,|\,\vec{0}_3]$ with row reduction, we find

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

We see that all of the columns of A are pivot columns, i.e., there are no free variables and hence no nontrivial solutions. We conclude that the columns of A are linearly independent.

Exercise 4.1.2. *For each matrix A, determine if the columns are linearly dependent or linearly independent. If dependent, find a linear dependence relation.*

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

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

Theorem 4.1.1 provides us with a standard approach to determining whether a given set of vectors is linearly independent or dependent. We can always create a matrix whose columns are the vectors in our set. We should be warned, however, that when determining linear dependence/independence of a collection of three or more vectors, it is generally not sufficient to consider smaller subsets of the vectors. For example, the reader should verify that each of the subsets of vectors

$$\{\vec{v}_1, \vec{v}_2\}, \quad \{\vec{v}_1, \vec{v}_3\} \quad \text{and} \quad \{\vec{v}_2, \vec{v}_3\}$$

from Example 4.1.1 are **linearly independent** despite the fact that the set of all three, $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, is linearly dependent.

While we cannot simply consider subsets of a collection of vectors when investigating their linear dependence/independence, we can say that a set of two or more vectors is linearly dependent if and only if at least one vector in the set can be written as a linear combination of the other vectors. Hence, if we can immediately recognize one of the vectors as a linear combination of one or more of the others, we can conclude that the set is linearly dependent without further investigation.

Example 4.1.3. *Determine whether the collection of vectors $\vec{v}_1 = \langle 1, -1, 1, -1 \rangle$, $\vec{v}_2 = \langle 2, 0, 3, 0 \rangle$, $\vec{v}_3 = \langle -3, 3, -3, 3 \rangle$ is linearly dependent or linearly independent. If linearly dependent, give a linear dependence relation.*

We can follow the procedure used in Example 4.1.1. However, we might simply observe that $\vec{v}_3$ is obtained from $\vec{v}_1$ by scalar multiplication.

$$\vec{v}_3 = \langle -3, 3, -3, 3 \rangle = -3\langle 1, -1, 1, -1 \rangle = -3\vec{v}_1.$$

Not only does this allow us to conclude that the vectors are linearly dependent, we can also rearrange this equation to write a linear dependence relation,

$$3\vec{v}_1 + \vec{v}_3 = \vec{0}_4.$$

Note that this is the same as writing

$$3\vec{v}_1 + 0\vec{v}_2 + \vec{v}_3 = \vec{0}_4.$$

To be a valid linear dependence relation, we require at least one of the coefficients to be nonzero. In this case, two of the three coefficients are nonzero.

There are cases for which linear dependence can be determined with very little effort. In particular, we have the following theorem.

Theorem 4.1.2. *Let $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ be a collection of k of vectors in R^m. If*

a. one of the vectors, say $\vec{v}_i = \vec{0}_m$, or if

b. $k > m$,

then the collection is linearly dependent.

The statement a. says that any set of vectors that includes the zero vector is necessarily linearly dependent. The proof of a. is left as an exercise (see Exercise 1). As for part b., note that this result says that if the number of vectors in our set is **larger** than the number of entries in each vector, the set is automatically linearly dependent. We can argue this using Corollary 2.4.1 to Theorem 2.4.1 from Chapter 2.

Proof. (of Theorem 4.1.2 part b.) Suppose that we have a set of k vectors $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$, each of which is in R^m, with $k > m$. Defining the matrix A such that $\text{Col}_i(A) = \vec{v}_i$,

$$\underbrace{\text{m}\left\{\begin{bmatrix} v_{11} & v_{21} & \cdots & v_{k1} \\ v_{12} & v_{22} & \cdots & v_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1m} & v_{2m} & \cdots & v_{km} \end{bmatrix}\right.}_{k>m},$$

the homogeneous system $A\vec{x} = \vec{0}_m$ will have more variables than equations—A has more columns than rows. Given that the system is homogeneous, it is necessarily consistent, and by Corollary 2.4.1, it must have infinitely many solutions. That is, there exists a nontrivial solution to $A\vec{x} = \vec{0}_m$, making the set of vectors linearly dependent. □

Exercise 4.1.3. *Without performing any computations explain why each of the following sets of vectors is linearly dependent.*

1. $\{\langle 1,0,0\rangle, \langle 0,1,0\rangle, \langle 0,0,0\rangle\}$
2. $\{\langle 1,2\rangle, \langle 2,1\rangle, \langle 3,0\rangle\}$
3. $\{\langle 1,0,0\rangle, \langle 0,1,0\rangle, \langle 0,0,1\rangle, \langle 1,-1,1\rangle\}$
4. $\{\langle 1,2,1\rangle, \langle -1,-2,-1\rangle, \langle 1,0,0\rangle\}$

Exercise 4.1.4. *For each set of vectors, determine if the set is linearly dependent or linearly independent. If dependent, find a linear dependence relation.*

1. $\{\langle 1,1,2\rangle, \langle 2,-1,0\rangle, \langle 1,-3,1\rangle\}$
2. $\{\langle 1,2,0,-3\rangle, \langle -2,-4,0,6\rangle, \langle 0,2,3,1\rangle, \langle 1,6,6,-1\rangle\}$
3. $\{\langle 0,4,-2,5\rangle, \langle 3,7,-5,-4\rangle, \langle 1,5,-3,2\rangle\}$
4. $\{\langle 3,1,0,-1\rangle, \langle 2,0,-2,8\rangle, \langle 3,1,5,4\rangle\}$

4.2 Subspaces of R^n

Since R^n is a set of objects (vectors), we can think of endless examples of subsets of R^n. For example,

- $\{\langle 1,2\rangle, \langle 3,-4\rangle, \langle 4,\pi\rangle\}$,
- $\{\langle 1,0\rangle, \langle 0,0\rangle\}$,
- $\{\langle k,k\rangle \mid k = 1,2,\ldots\}$, and
- $\text{Span}\{\langle 1,0\rangle\}$

are all subsets of R^2. There is something fundamentally different about the last example, Span$\{\langle 1,0\rangle\}$, and not just because it includes infinitely many elements (the third example also includes an infinitude of elements). What makes Span$\{\langle 1,0\rangle\}$ distinct from these other subsets is that it has a similar structure to R^1, while being a subset[3] of R^2. Most notably, when we take any elements of the set Span$\{\langle 1,0\rangle\}$ and perform the operations of vector addition or scalar multiplication, the resulting vector is also an element of Span$\{\langle 1,0\rangle\}$. This is one of the key properties of algebra that mathematicians associate with the word *space*. The subset Span$\{\langle 1,0\rangle\}$ is an example of a **subspace** of R^2.

Definition 4.2.1. *A subset, S, of R^n is a* **subspace** *of R^n provided*

i. S is nonempty,

ii. for any pair of vectors $\vec{u}$ and $\vec{v}$ in S, $\vec{u}+\vec{v}$ is in S, and

iii. for any vector $\vec{u}$ in S and scalar c in R, $c\vec{u}$ is in S.

The properties ii. and iii. are referred to as being "*closed* with respect to" (or "closed under") the indicated operation. That is, we can state property ii. by saying:

"A subspace of R^n is closed with respect to vector addition."

We can similarly state property iii. by saying that a subspace of R^n is closed under scalar multiplication.

Example 4.2.1. *Let $K = \{\langle a,0,b\rangle \in R^3 \mid -\infty < a,b < \infty\}$. Show that K is a subspace of R^3.*

Note that we can characterize K as the subset of R^3 of vectors having second entry zero. We will show that K satisfies all the properties of Definition 4.2.1. First, we see that K is not empty—for example, $\langle 0,0,0\rangle$ is in K. To show that property ii. holds, let's consider any two vectors $\vec{u}$ and $\vec{v}$ in K, say $\vec{u} = \langle a_1,0,b_1\rangle$ and $\vec{v} = \langle a_2,0,b_2\rangle$ where $a_1, b_1, a_2,$ and b_2 are any real numbers. Note that

$$\vec{u}+\vec{v} = \langle a_1,0,b_1\rangle + \langle a_2,0,b_2\rangle = \langle a_1+a_2, 0+0, b_1+b_2\rangle = \langle a_1+a_2, 0, b_1+b_2\rangle.$$

[3] To be clear, we are not equating Span$\{\langle 1,0\rangle\}$ with R^1. We are just noting that it has a similar structure in that it contains all the vectors obtained by doing arithmetic on its elements.

We see that $\vec{u}+\vec{v}$ has the required property to be an element of K—i.e., its second entry is zero. To show that property iii. holds, we let c be any scalar, and note that

$$c\vec{u} = c\langle a_1, 0, b_1\rangle = \langle ca_1, c(0), cb_1\rangle = \langle ca_1, 0, cb_1\rangle.$$

Here too, we see that $c\vec{u}$ has the property necessary to be an element of K. Since K is a nonempty subset of R^3 that is closed under vector addition and scalar multiplication, we can conclude that K is a subspace of R^3. We might recognize $K = \text{Span}\{\vec{e}_1, \vec{e}_3\}$ from Example 1.3.4.

Example 4.2.2. *Consider the subset F of R^2 given by*

$$F = \left\{\langle a, b\rangle \in R^2 \,|\, a, b \in R,\ a \geq 0\right\}.$$

Determine whether F is a subspace of R^2.

We can characterize F as the subset of vectors in R^2 having nonnegative first entry. We can think of F geometrically as the collection of all vectors whose standard representation has terminal point on or to the right of the y-axis—this is sometimes called the right half plane. *The subset F is clearly nonempty—for example the vector $\langle 0,0\rangle$ is in F. Moreover, F is closed under vector addition. Suppose $\vec{u} = \langle a_1, b_1\rangle$ and $\vec{v} = \langle a_2, b_2\rangle$ are in F. This requires $a_1 \geq 0$ and $a_2 \geq 0$ with b_1 and b_2 any real numbers. Note that*

$$\vec{u}+\vec{v} = \langle a_1, b_1\rangle + \langle a_2, b_2\rangle = \langle a_1 + a_2, b_1 + b_2\rangle.$$

Since $a_1 + a_2 \geq 0 + 0 = 0$, we see that $\vec{u}+\vec{v}$ is in F. However, we cannot conclude that F is a subspace. In fact, F is not closed under scalar multiplication. For example, the vector $\langle 1, 0\rangle$ is in F, but the vector

$$-1\langle 1, 0\rangle = \langle -1, 0\rangle$$

is **not** *in F, since its first entry is not nonnegative. We conclude that F is not a subspace of R^2.*

In the definition of subspace, some authors replace the condition that S is nonempty with the condition that S must contain the zero vector. We'll state this as a theorem here leaving the proof as an exercise (see Exercise 3).

Theorem 4.2.1. *If S is a subspace of R^n, then $\vec{0}_n$ is an element of S.*

This can be used as a quick test to rule out the possibility that a subset is a subspace—that is, if it doesn't contain the zero vector it is not a subspace. Be careful though. As Example 4.2.2 illustrates, the converse of Theorem 4.2.1 is not true. A subset may well contain the zero vector but still fail to be a subspace.

Exercise 4.2.1. *Determine whether each subset is a subspace of R^n for the indicated value of n.*

1. $S = \{\langle 0, a\rangle \in R^2 \mid a \in R\}$

2. $T = \{\langle 1, a\rangle \in R^2 \mid a \in R\}$

3. $Q = \{\langle 0, 0, 0\rangle\}$ *in* R^3

4. $P = \{\langle k, k\rangle \in R^2 \mid k = 1, 2, \ldots\}$

5. $L = \{\langle k, k\rangle \in R^2 \mid k \in R\}$

The following theorem follows directly from the definition of Span (Definition 1.3.2).

Theorem 4.2.2. *If $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is any nonempty subset of vectors in R^n, then the set* $\mathrm{Span}(S)$ *is a subspace of R^n.*

We can prove Theorem 4.2.2 by demonstrating that a span necessarily satisfies the three subspace criteria.

Proof. First, suppose $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is nonempty. Then $\mathrm{Span}(S)$ will contain the elements of S making it also nonempty. Next, suppose $\vec{y}$ and $\vec{z}$ are any elements of $\mathrm{Span}(S)$. Then we can write $\vec{y}$ and $\vec{z}$ as linear combinations of elements of S,

$$\begin{aligned} \vec{y} &= a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_k\vec{v}_k \\ \vec{z} &= b_1\vec{v}_1 + b_2\vec{v}_2 + \cdots + b_k\vec{v}_k \end{aligned}$$

for some sets of weights a_i, b_i, $i = 1, \ldots, k$. Then note that the sum,

$$\vec{y} + \vec{z} = (a_1 + b_1)\vec{v}_1 + (a_2 + b_2)\vec{v}_2 + \cdots + (a_k + b_k)\vec{v}_k,$$

is a linear combination of the elements of S. That is, $\vec{y} + \vec{z}$ is contained in Span(S), and Span(S) is closed under vector addition. Similarly, if c is any scalar, then the scalar product

$$c\vec{y} = ca_1\vec{v}_1 + ca_2\vec{v}_2 + \cdots + ca_k\vec{v}_k,$$

is also a linear combination of the elements of S. Thus Span(S) is also closed under scalar multiplication. We conclude that Span(S) is a subspace of R^n, which completes the proof. □

This provides us with an alternative approach to investigating whether a given subset is in fact a subspace. If we can find a spanning set for a given subset of R^n, Theorem 4.2.2 says we can conclude that our subset is a subspace.

Example 4.2.3. *Show that the set $Q = \{\langle 0, a, b, a+b\rangle \in R^4 \mid a, b \in R\}$ is a subspace of R^4.*

We can generate a spanning set by starting with a representative element of the set Q and decomposing it as a linear combination of fixed vectors with parameter coefficients—by fixed *vectors, we mean vectors whose entries are specific numbers, not variables or parameters. Fortunately, there is a simple, intuitive approach. We can simply think of factoring a vector according to the values that can vary, in this example, a and b. Note that for an arbitrary element of Q,*

$$\langle 0, a, b, a+b\rangle = \langle 0, a, 0, a\rangle + \langle 0, 0, b, b\rangle = a\langle 0, 1, 0, 1\rangle + b\langle 0, 0, 1, 1\rangle.$$

So the elements of Q are linear combinations of the pair $\langle 0, 1, 0, 1\rangle$ and $\langle 0, 0, 1, 1\rangle$, that is

$$Q = \text{Span}\{\langle 0, 1, 0, 1\rangle, \langle 0, 0, 1, 1\rangle\}.$$

By Theorem 4.2.2 we conclude that Q is a subspace of R^4.

Exercise 4.2.2. *Find a spanning set for each subspace of R^n.*

- $Q = \{\langle 0, a\rangle \in R^2 \mid a \in R\}$
- $P = \{\langle a, a, b\rangle \in R^3 \mid a, b \in R\}$
- $T = \{\langle a, b, c, a+b+c\rangle \in R^4 \mid a, b, c \in R\}$

4.2.1 The Fundamental Subspaces of a Matrix

In Section 3.3, we associated two sets of vectors with an $m \times n$ matrix A. These were the row vectors $\{\text{Row}_1(A), \ldots, \text{Row}_m(A)\}$, which are vectors in R^n, and the column vectors $\{\text{Col}_1(A), \ldots, \text{Col}_n(A)\}$, which are vectors in R^m. This provides us with two subspaces, one of R^n and one of R^m, that we can associated with a given matrix.

Definition 4.2.2. *Let A be an $m \times n$ matrix. The subspace of R^n spanned by the row vectors of A, denoted*

$$\mathcal{RS}(A) = \text{Span}\{\text{Row}_1(A), \ldots, \text{Row}_m(A)\},$$

is called the **row space of** A.

Definition 4.2.3. *Let A be an $m \times n$ matrix. The subspace of R^m spanned by the column vectors of A, denoted*

$$\mathcal{CS}(A) = \text{Span}\{\text{Col}_1(A), \ldots, \text{Col}_n(A)\},$$

is called the **column space of** A.

The connection between $\mathcal{RS}(A)$ and systems of equations or R^n in general is not obvious at present. But thinking back to the introduction of this chapter, we see that the column space of a matrix is of immediate interest when considering systems of linear equations. Now we can say that a product $A\vec{x}$ for $\vec{x}$ in R^n is an element of $\mathcal{CS}(A)$. So $\mathcal{CS}(A)$ is the set of all $\vec{y}$ in R^n such that the system $A\vec{x} = \vec{y}$ is consistent.

Example 4.2.4. *Let* $A = \begin{bmatrix} 0 & 3 & 1 \\ 4 & 7 & 5 \\ -2 & -5 & -3 \\ 5 & -4 & 2 \end{bmatrix}$. *Identify a spanning set for $\mathcal{RS}(A)$ and a spanning set for $\mathcal{CS}(A)$.*

We can simply apply the definitions since the spanning sets can be the row vector or the column vectors, as appropriate. A spanning set for $\mathcal{RS}(A)$ is

$$\{\langle 0, 3, 1\rangle, \langle 4, 7, 5\rangle, \langle -2, -5, -3\rangle, \langle 5, -4, 2\rangle\}.$$

A spanning set for $\mathcal{CS}(A)$ is

$$\{\langle 0, 4, -2, 5\rangle, \langle 3, 7, -5, -4\rangle, \langle 1, 5, -3, 2\rangle\}.$$

It's worth noting that the vectors in the spanning set for $\mathcal{RS}(A)$ are elements of R^3, whereas the vectors in the spanning set for $\mathcal{CS}(A)$ are elements of R^4.

Exercise 4.2.3. *Consider the set of vectors*

$$T = \{\langle 1,0,1,1\rangle, \langle -2,3,0,8\rangle, \langle 4,4,5,2\rangle\}.$$

1. *Find a matrix A having row space $\mathcal{RS}(A) = \text{Span}(T)$.*

2. *Find a matrix A having columns space $\mathcal{CS}(A) = \text{Span}(T)$.*

The row space and column spaces are two of four subspaces associated with a matrix. Taken together these are usually referred to as **fundamental subspaces** of the matrix. A third can be defined by consideration of the homogeneous system $A\vec{x} = \vec{0}_m$. We will call the set of all solutions of this homogeneous equation the **null space** of the matrix A.

Definition 4.2.4. *Let A be an $m \times n$ matrix. The* **null space** *of A, denoted $\mathcal{N}(A)$, is the set of all solutions of the homogeneous equation $A\vec{x} = \vec{0}_m$. That is,*

$$\mathcal{N}(A) = \{\vec{x} \in R^n \mid A\vec{x} = \vec{0}_m\}.$$

Unlike the row and column spaces, the null space is not defined in terms of a spanning set. Nevertheless, the null space of a matrix is a subspace of R^n.

Theorem 4.2.3. *If A is an $m \times n$ matrix, then $\mathcal{N}(A)$ is a subspace of R^n.*

Proof. That $\mathcal{N}(A)$ is a subset of R^n follows from the definition of the product $A\vec{x}$. We can prove Theorem 4.2.3 by demonstrating that $\mathcal{N}(A)$ satisfies the properties of a subspace. First, the equation always permits the trivial solution, hence $\vec{0}_n \in \mathcal{N}(A)$ so that $\mathcal{N}(A)$ is nonempty. Next, suppose $\vec{u}$ and $\vec{v}$ are any elements of $\mathcal{N}(A)$—meaning $A\vec{u} = \vec{0}_m$ and $A\vec{v} = \vec{0}_m$, and let c be any scalar. Using the distributive property of the matrix-vector product

$$A(\vec{u}+\vec{v}) = A\vec{u} + A\vec{v} = \vec{0}_m + \vec{0}_m = \vec{0}_m, \quad \text{i.e., } \vec{u}+\vec{v} \in \mathcal{N}(A),$$

and using the fact that we can factor scalars

$$A(c\vec{u}) = cA\vec{u} = c\vec{0}_m = \vec{0}_m, \quad \text{i.e., } c\vec{u} \in \mathcal{N}(A).$$

Thus $\mathcal{N}(A)$ is closed under vector addition and scalar multiplication. This establishes that $\mathcal{N}(A)$ is a subspace of R^n completing the proof of Theorem 4.2.3. □

Example 4.2.5. *Let* $A = \begin{bmatrix} 3 & 3 & 3 \\ -1 & -2 & 1 \\ -5 & -6 & -3 \end{bmatrix}$. *Find a spanning set for* $\mathcal{N}(A)$.

We need to find a representation for solutions to the homogeneous equation $A\vec{x} = \vec{0}_3$. *We can use row reduction on* $\left[A \mid \vec{0}_3\right]$.

$$\left[\begin{array}{rrr|r} 3 & 3 & 3 & 0 \\ -1 & -2 & 1 & 0 \\ -5 & -6 & -3 & 0 \end{array}\right] \xrightarrow{rref} \left[\begin{array}{rrr|r} 1 & 0 & 3 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

From the rref, we see that a solution $\vec{x} = \langle x_1, x_2, x_3 \rangle$ *will have*

$$x_1 = -3x_3, \quad x_2 = 2x_3, \quad \text{and} \quad x_3 \text{ is free.}$$

This means that any solution can be written as $\vec{x} = \langle -3x_3, 2x_3, x_3 \rangle$ *for some real number* x_3. *To find a spanning set, we can decompose such a vector (using the same process we used in Example 4.2.3 and Exercise 4.2.2) to write it as a linear combination*

$$\vec{x} = \langle -3x_3, 2x_3, x_3 \rangle = x_3 \langle -3, 2, 1 \rangle.$$

So a spanning set for $\mathcal{N}(A)$ *is the single element set* $\{\langle -3, 2, 1 \rangle\}$.

Exercise 4.2.4. *Find a spanning set for the null space* $\mathcal{N}(A)$ *where*

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

Note that $\mathcal{RS}(A)$ and $\mathcal{N}(A)$ are both subspaces of R^n whereas $\mathcal{CS}(A)$ is a subspace of R^m. An interesting interpretation of $\mathcal{RS}(A)$ stems from the observation that if $\vec{u}$ is any vector in $\mathcal{RS}(A)$ and $\vec{x}$ is any vector in $\mathcal{N}(A)$, then it is necessarily the case that

$$\vec{u} \cdot \vec{x} = 0.$$

To see that this is true, we recall from Section 3.5 that one of the interpretations of the product $A\vec{x}$ is as the vector of dot products of the rows of A with $\vec{x}$,

$$A\vec{x} = \langle \text{Row}_1(A) \cdot \vec{x}, \text{Row}_2(A) \cdot \vec{x}, \ldots, \text{Row}_m(A) \cdot \vec{x} \rangle.$$

Hence, if $\vec{x}$ is in $\mathcal{N}(A)$, each entry in $A\vec{x}$ must be zero—meaning each $\text{Row}_i(A) \cdot \vec{x} = 0$. If we take any element of $\mathcal{RS}(A)$, say $\vec{u}$, then we can write this vector as a linear combination of the rows of A,

$$\vec{u} = c_1 \,\text{Row}_1(A) + c_2 \,\text{Row}_2(A) + \cdots + c_m \,\text{Row}_m(A),$$

for some set of weights, $c_1, \ldots, c_m$. Let $\vec{x}$ be any element of $\mathcal{N}(A)$. We can form the dot product $\vec{x} \cdot \vec{u}$, and using the algebraic properties of the dot product, we get

$$\begin{aligned} \vec{x} \cdot \vec{u} &= \vec{x} \cdot \Big(c_1 \,\text{Row}_1(A) + c_2 \,\text{Row}_2(A) + \cdots + c_m \,\text{Row}_m(A)\Big) \\ &= c_1\vec{x} \cdot \text{Row}_1(A) + c_2\vec{x} \cdot \text{Row}_2(A) + \cdots + c_m\vec{x} \cdot \text{Row}_m(A) \\ &= c_1(0) + c_2(0) + \cdots + c_m(0) \\ &= 0. \end{aligned}$$

We conclude that every vector in $\mathcal{RS}(A)$ is orthogonal to every vector in $\mathcal{N}(A)$. Due to this property, that every vector in one set is orthogonal to every vector in the other set, we refer to these as **orthogonal complements**. That is, for any matrix A, $\mathcal{RS}(A)$ is the orthogonal complement of $\mathcal{N}(A)$—and vice versa. In fact, the fourth fundamental subspace associated with a matrix A is the orthogonal complement of the column space.

Recall that in Section 3.4, we defined the transpose of an $m \times n$ matrix A to be the $n \times m$ matrix A^T such that

$$\text{Row}_i(A^T) = \text{Col}_i(A), \quad \text{for } i = 1, \ldots, n.$$

Similarly,

$$\text{Col}_i(A^T) = \text{Row}_i(A), \quad \text{for } i = 1, \ldots, m.$$

Since the rows and columns of A and A^T are swapped, we can immediately see that

$$\mathcal{RS}(A^T) = \mathcal{CS}(A), \quad \text{and} \quad \mathcal{CS}(A^T) = \mathcal{RS}(A).$$

The fourth fundamental subspace associated with a matrix A, the orthogonal complement of its column space, is the null space of A^T, $\mathcal{N}(A^T)$. We can think of $\mathcal{N}(A^T)$ as being defined by

$$\mathcal{N}(A^T) = \{\vec{x} \in R^m \,|\, \vec{x} \cdot \vec{y} = 0, \text{ for every } \vec{y} \in \mathcal{CS}(A)\}\,. \tag{4.4}$$

This definition highlights its property of being the orthogonal complement of $\mathcal{CS}(A)$. Or, we can define $\mathcal{N}(A^T)$ by

$$\mathcal{N}(A^T) = \left\{ \vec{x} \in R^m \mid A^T\vec{x} = \vec{0}_n \right\}, \tag{4.5}$$

highlighting its relationship to a specific homogeneous equation. Of course, the sets defined in equations (4.4) and (4.5) are the same set, and like $\mathcal{CS}(A)$, $\mathcal{N}(A^T)$ is a subspace of R^m.

Example 4.2.6. *Let* $A = \begin{bmatrix} 3 & 3 & 3 \\ -1 & -2 & 1 \\ -5 & -6 & -3 \end{bmatrix}$. *Find a spanning set for the orthogonal complement of* $\mathcal{CS}(A)$.

We know that the orthogonal complement of $\mathcal{CS}(A)$ *is the null space of* A^T, *i.e.* $\mathcal{N}(A^T)$. *So we can restate the problem as finding a spanning set for* $\mathcal{N}(A^T)$, *which in turn means we need to find a representation for solutions of the homogeneous equation* $A^T\vec{x} = \vec{0}_3$. *Using row reduction on the augmented matrix* $\left[A^T \mid \vec{0}_3\right]$.

$$\left[\begin{array}{ccc|c} 3 & -1 & -5 & 0 \\ 3 & -2 & -6 & 0 \\ 3 & 1 & -3 & 0 \end{array}\right] \quad \xrightarrow{rref} \quad \left[\begin{array}{ccc|c} 1 & 0 & -4/3 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

We see that a solution $\vec{x} = \langle x_1, x_2, x_3 \rangle$ *of* $A^T\vec{x} = \vec{0}_3$ *will have entries*

$$x_1 = \frac{4}{3}x_3, \quad x_2 = -x_3, \quad \text{with } x_3 \text{ free.}$$

Decomposing such a vector,

$$\vec{x} = \left\langle \frac{4}{3}x_3, -x_3, x_3 \right\rangle = x_3 \left\langle \frac{4}{3}, -1, 1 \right\rangle.$$

A spanning set for the orthogonal complement of $\mathcal{CS}(A)$ *(a.k.a.* $\mathcal{N}(A^T)$*) is*

$$\left\{ \left\langle \frac{4}{3}, -1, 1 \right\rangle \right\}.$$

Example 4.2.7. *Find a spanning set for each of the four fundamental subspaces of the matrix*

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

For the row and columns spaces, we can simply use the row and column vectors, respectively. That is,

$$\mathcal{RS}(A) = \text{Span}\{\langle 1,-2,5,4\rangle, \langle 2,-4,1,-1\rangle\}, \quad \textit{and}$$

$$\mathcal{CS}(A) = \text{Span}\{\langle 1,2\rangle, \langle -2,-4\rangle, \langle 5,1\rangle, \langle 4,-1\rangle\}.$$

For the two null spaces, we will have to consider homogeneous equations and do row reduction. To characterize $\mathcal{N}(A)$*, we row reduce* $\left[A \,|\, \vec{0}_2\right]$.

$$\left[\begin{array}{cccc|c} 1 & -2 & 5 & 4 & 0 \\ 2 & -4 & 1 & -1 & 0 \end{array}\right] \quad \xrightarrow{rref} \quad \left[\begin{array}{cccc|c} 1 & -2 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{array}\right].$$

A solution $\vec{x}$ *to* $A\vec{x} = \vec{0}$ *will have entries*

$$x_1 = 2x_2 + x_4, \quad x_3 = -x_4, \quad \textit{with } x_2, x_4 \textit{ free.}$$

Hence

$$\vec{x} = \langle 2x_2 + x_4, x_2, -x_4, x_4\rangle = x_2\langle 2,1,0,0\rangle + x_4\langle 1,0,-1,1\rangle.$$

A spanning set for $\mathcal{N}(A)$ *is*

$$\mathcal{N}(A) = \text{Span}\left\{\langle 2,1,0,0\rangle, \langle 1,0,-1,1\rangle\right\}.$$

Playing a similar game to find a spanning set for $\mathcal{N}(A^T)$*, we set up and reduce* $\left[A^T \,|\, \vec{0}_4\right]$

$$\left[\begin{array}{cc|c} 1 & 2 & 0 \\ -2 & -4 & 0 \\ 5 & 1 & 0 \\ 4 & -1 & 0 \end{array}\right] \quad \xrightarrow{rref} \quad \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

We see that if $\vec{x} = \langle x_1, x_2\rangle$ *is in* $\mathcal{N}(A^T)$*, then* $x_1 = x_2 = 0$*. That is,*

$$\mathcal{N}(A^T) = \{\langle 0,0\rangle\}.$$

We can write $\mathcal{N}(A^T) = \text{Span}\{\langle 0,0\rangle\}$.

Exercise 4.2.5. *For each matrix A, find a spanning set for each of the four fundamental subspaces of A.*

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

2. $A = \begin{bmatrix} 1 & 3 & 1 & 0 & 11 \\ -1 & 1 & 4 & 1 & 0 \\ -2 & 0 & 3 & -1 & -3 \end{bmatrix}$

3. $A = \begin{bmatrix} 4 & 8 & -3 & 1 \\ -2 & -4 & 5 & -11 \\ 3 & 6 & 1 & -9 \end{bmatrix}$

4.3 Bases

We can think of a spanning set as a set of building blocks that, when combined with the operations of vector addition and scalar multiplication, can be used to generate an entire subspace. But not all spanning sets are created equally. Consider the vector space R^2. In Example 1.3.3, we saw that the set $\{\langle 1,0\rangle, \langle 0,1\rangle\}$ is a spanning set for R^2. In some sense, it seems like an obvious choice of a spanning set, but it's certainly not the only one. For example, we also saw that $R^2 = \text{Span}\{\langle 1,1\rangle, \langle 1,-1\rangle\}$ and $R^2 = \text{Span}\{\langle 2,1\rangle, \langle 1,2\rangle\}$ (see Exercise 8 at the end of Chapter 1). Similarly,

$$\langle x_1, x_2\rangle = (x_1 - x_2 + 1)\langle 1,0\rangle + (x_2 - 1)\langle 1,1\rangle + \langle 0,1\rangle,$$

so we also have $R^2 = \text{Span}\{\langle 1,0\rangle, \langle 1,1\rangle, \langle 0,1\rangle\}$. At the same time, the set $\{\langle 1,0\rangle\}$ does not span R^2 since every linear combination of $\langle 1,0\rangle$ will necessarily have second component zero. While it's not possible to talk about *the* (i.e., a unique) spanning set for a given subspace of R^n, we can ask whether there is a minimum amount of information (number of vectors) necessary to construct all of the vectors in a subspace. We can also ask whether a given spanning set includes superfluous information—such as the set $\{\langle 1,0\rangle, \langle 1,1\rangle, \langle 0,1\rangle\}$ whose proper subset $\{\langle 1,0\rangle, \langle 0,1\rangle\}$ spans R^2.

Example 4.3.1. *Suppose* $\vec{u}_1, \vec{u}_2$ *and* $\vec{u}_3$ *are vectors in* R^n *with*

$$\vec{u}_3 = \vec{u}_1 + 3\vec{u}_2.$$

Show that $\text{Span}\{\vec{u}_1, \vec{u}_2, \vec{u}_3\} = \text{Span}\{\vec{u}_1, \vec{u}_2\}$.

First, we will note that if $\vec{x}$ *is any vector in* $\text{Span}\{\vec{u}_1, \vec{u}_2\}$, *then there are scalars* c_1 *and* c_2 *such that*

$$\vec{x} = c_1\vec{u}_1 + c_2\vec{u}_2.$$

This is equivalent to

$$\vec{x} = c_1\vec{u}_1 + c_2\vec{u}_2 + 0\vec{u}_3,$$

so $\vec{x}$ *is necessarily in* $\text{Span}\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$. *We need to show that every vector in* $\text{Span}\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ *is in* $\text{Span}\{\vec{u}_1, \vec{u}_2\}$. *To that end, suppose* $\vec{x}$ *is in* $\text{Span}\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$, *so that*

$$\vec{x} = c_1\vec{u}_1 + c_2\vec{u}_2 + c_3\vec{u}_3,$$

for some scalars c_1, c_2 *and* c_3. *Since* $\vec{u}_3 = \vec{u}_1 + 3\vec{u}_2$, *we an rearrange the above to get*

$$\vec{x} = c_1\vec{u}_1 + c_2\vec{u}_2 + c_3(\vec{u}_1 + 3\vec{u}_2) = (c_1 + c_3)\vec{u}_1 + (c_2 + 3c_3)\vec{u}_2.$$

That is, $\vec{x}$ *is a linear combination of the pair* $\vec{u}_1$ *and* $\vec{u}_2$, *and we can conclude that* $\text{Span}\{\vec{u}_1, \vec{u}_2, \vec{u}_3\} = \text{Span}\{\vec{u}_1, \vec{u}_2\}$.

Example 4.3.1 illustrates that if a spanning set is linearly dependent (which is the case since $\vec{u}_3$ is a linear combination of $\vec{u}_1$ and $\vec{u}_2$), we may be able to discard one or more elements without losing part of the subspace generated by the set. To generate a given subspace, we require the vectors in a set to span the subspace. Linear independence is the key property that is needed to ensure that a spanning set does not include superfluous information. A minimal spanning set is called a **basis**. We have the following definition.

Definition 4.3.1. *Let* S *be a subspace of* R^n, *and let* $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_k\}$ *be a subset of vectors in* S. $\mathcal{B}$ *is a* **basis** *of* S *provided*

- $\mathcal{B}$ *spans* S, *and*

- *$\mathcal{B}$ is linearly independent.*

So a basis is a linearly independent spanning set. A basis $\mathcal{B}$ contains all of the information needed to construct every vector in a subspace, and a basis is minimal in the sense that the subspace is not spanned by any proper subset of $\mathcal{B}$.

Example 4.3.2. *The column vectors of the $n\times n$ identity matrix form a basis for R^n. Recall that we used the notation $\vec{e}_i$ to denote such a vector having 1 in the i^{th} position and zero everywhere else. The set $\mathcal{E} = \{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$ of standard unit vectors in R^n is called the* **standard basis** *or the* **elementary basis** *of R^n. Note that any vector $\vec{x} = \langle x_1, x_2, \dots, x_n\rangle$ can be expressed as a linear combination in the rather obvious way,*

$$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n,$$

so $\mathcal{E}$ spans R^n. Moreover, as the columns of the $n \times n$ identity matrix (which is its own inverse), we know that the set $\mathcal{E}$ is linearly independent.

Example 4.3.3. *Show that the set $\mathcal{B} = \{\langle 2, 1\rangle, \langle 1, 2\rangle\}$ is a basis of R^2.*

First, we note that in Exercise 8 at the end of Chapter 1, we showed that this set spans[4] *R^2. So to conclude that $\mathcal{B}$ is a basis, we have to show that the set $\{\langle 2, 1\rangle, \langle 1, 2\rangle\}$ is linearly independent. There are various approaches we can use. Let's define a matrix B having the vectors as columns and consider the homogeneous equation $B\vec{x} = \vec{0}_2$. Using row reduction, note that*

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

We see that B has two pivot columns so that there are no nontrivial solutions to $B\vec{x} = \vec{0}_2$. Applying Theorem 4.1.1, it follows that the set $\mathcal{B}$ is linearly independent. Hence $\mathcal{B}$ is a basis for R^2.

---

[4] To refresh our memories,

$$\langle x_1, x_2\rangle = \left(\frac{2x_2 - x_1}{3}\right)\langle 1, 2\rangle + \left(\frac{2x_1 - x_2}{3}\right)\langle 2, 1\rangle.$$

Example 4.3.4. *Determine whether the set*

$$\mathcal{S} = \{\langle 1,0,0\rangle, \langle 0,1,0\rangle, \langle 0,0,2\rangle, \langle 1,1,1\rangle\}$$

is a basis for R^3.

We can easily write any vector $\langle x_1, x_2, x_3\rangle$ *as a linear combination of the vectors* $\mathcal{S}$. *Note for example that*

$$\langle x_1, x_2, x_3\rangle = x_1\langle 1,0,0\rangle + x_2\langle 0,1,0\rangle + \frac{x_3}{2}\langle 0,0,2\rangle.$$

So $\mathcal{S}$ *certainly spans* R^3. *However, the set* $\mathcal{S}$ *contains four elements, each with three entries. By Theorem 4.1.2,* $\mathcal{S}$ *is linearly dependent. Hence, it is not a basis for* R^3.

Exercise 4.3.1. *Suppose* S *is a subspace of* R^n *for some* $n \geq 2$, *and let* $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_k\}$ *be a basis for* S. *Explain why the number of vectors,* k, *in the basis* $\mathcal{B}$ *must be less than or equal to* n.

Example 4.3.5. *Find a basis for the subspace* $Q = \{\langle 0, a, b, a+b\rangle \in R^4 \mid a, b \in R\}$ *of* R^4.

In Example 4.2.3, we decomposed an arbitrary element of Q *to find a spanning set. The process that we used there has the added benefit of producing a spanning set that is actually a basis. We wrote an element of* Q *as*

$$\langle 0, a, b, a+b\rangle = \langle 0, a, 0, a\rangle + \langle 0, 0, b, b\rangle = a\langle 0,1,0,1\rangle + b\langle 0,0,1,1\rangle,$$

giving us the spanning set $\{\langle 0,1,0,1\rangle, \langle 0,0,1,1\rangle\}$. *Each of these vectors has some entry in which one vector has a 1 where the other has a zero. The second entry of* $\langle 0,1,0,1\rangle$ *is* 1 *and the third is* 0, *where as the second entry of* $\langle 0,0,1,1\rangle$ *is* 0 *and the third entry is* 1. *This guarantees linear independence. Note that*

$$c_1\langle 0,1,0,1\rangle + c_2\langle 0,0,1,1\rangle = \langle 0,0,0,0\rangle \implies \langle 0, c_1, c_2, c_1+c_2\rangle = \langle 0,0,0,0\rangle.$$

In particular, the second and third entries give

$$\begin{array}{rcl} c_1 & = & 0 \\ c_2 & = & 0 \end{array}.$$

This shows that the set $\{\langle 0,1,0,1\rangle, \langle 0,0,1,1\rangle\}$ *is linearly independent, and hence is a basis for* Q.

Exercise 4.3.2. *Find a basis for each subspace of R^n. You may wish to start with the spanning sets you found for these same subspaces in Exercise 4.2.2, but you should demonstrate that the set you claim is a basis is both a spanning set and is linearly independent.*

- $Q = \{\langle 0, a\rangle \in R^2 \mid a \in R\}$
- $P = \{\langle a, a, b\rangle \in R^3 \mid a, b \in R\}$
- $T = \{\langle a, b, c, a+b+c\rangle \in R^4 \mid a, b, c \in R\}$

Example 4.3.6. *Find a basis for the null space, $\mathcal{N}(A)$, of the matrix*

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

We recall that $\mathcal{N}(A)$ is set of all solutions of the homogeneous equation $A\vec{x} = \vec{0}_3$. Since A has size 3×4, $\mathcal{N}(A)$ will be a subspace of R^4. Fortunately, the process we've used to find a spanning set for $\mathcal{N}(A)$ will automatically generate a basis. To see this, let's first solve the homogeneous equation by using row reduction on $\left[A \mid \vec{0}_3\right]$. We have

$$\left[\begin{array}{rrrr|r} 1 & 3 & -1 & 5 & 0 \\ -1 & 1 & -3 & 3 & 0 \\ 2 & -2 & 6 & -6 & 0 \end{array}\right] \xrightarrow{rref} \left[\begin{array}{rrrr|r} 1 & 0 & 2 & -1 & 0 \\ 0 & 1 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Now, we can see that a solution $\vec{x} = \langle x_1, x_2, x_3, x_4\rangle$ will have two basic variables, x_1 and x_2, and two free variables, x_3 and x_4. We can write the solution

$$\begin{array}{rcl} x_1 & = & -2s + t \\ x_2 & = & s - 2t \\ x_3 & = & s \\ x_4 & = & t \end{array},$$

where $-\infty < s, t < \infty$. So a representative solution will be a vector of the form

$$\vec{x} = \langle -2s + t, s - 2t, s, t\rangle.$$

As we've done before, we can decompose this vector as a linear combination by factoring out the parameters s and t.

$$\begin{aligned}\vec{x} &= \langle -2s+t, s-2t, s, t\rangle \\ &= \langle -2s, s, s, 0\rangle + \langle t, -2t, 0, t\rangle \\ &= s\langle -2, 1, 1, 0\rangle + t\langle 1, -2, 0, 1\rangle\end{aligned}$$

So an element $\vec{x}$ of $\mathcal{N}(A)$ will be a linear combination of the vectors $\langle -2, 1, 1, 0\rangle$ and $\langle 1, -2, 0, 1\rangle$. If we focus on the third and fourth entries—because the free variables were x_3 and x_4, we'll see that the pair of vectors $\{\langle -2, 1, 1, 0\rangle, \langle 1, -2, 0, 1\rangle\}$ is necessarily linearly independent! If we set up the linear combination

$$c_1\langle -2, 1, 1, 0\rangle + c_2\langle 1, -2, 0, 1\rangle = \langle 0, 0, 0, 0\rangle,$$

the third entry gives $c_1 = 0$ and the fourth entry gives $c_2 = 0$, confirming the linear independence. If we obtain a spanning set for a null space by decomposing solutions in this fashion (separating out the free variables), we're guaranteed to produce a basis! Our solution, a basis for $\mathcal{N}(A)$, is the set

$$\{\langle -2, 1, 1, 0\rangle, \langle 1, -2, 0, 1\rangle\}.$$

Exercise 4.3.3. *Find a basis for $\mathcal{N}(A)$ and for $\mathcal{N}(A^T)$ for each matrix A.*

1. $A = \begin{bmatrix} 1 & 3 & 0 & 1 \\ 2 & 6 & 0 & 2 \\ 3 & 9 & 1 & 2 \end{bmatrix}$

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

4.3.1 Coordinate Vectors

A defining characteristic of a basis, as opposed to a spanning set, is that if $\mathcal{B}$ is a basis for a subspace S of R^n, and $\vec{x}$ is any element of S, then there is exactly one representation of $\vec{x}$ (i.e., exactly one set of coefficients) as a linear combination of the basis elements. To demonstrate this, suppose $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_k\}$ is an ordered[5] basis for some subspace S and suppose an

[5]By ordered, we simply mean that we have assigned an order to the vectors in the set and have numbered the vectors accordingly.

element $\vec{x}$ of S can be written in two, potentially different, ways:

$$\vec{x} = a_1\vec{b}_1 + a_2\vec{b}_2 + \cdots + a_k\vec{b}_k \tag{4.6}$$
$$\vec{x} = c_1\vec{b}_1 + c_2\vec{b}_2 + \cdots + c_k\vec{b}_k \tag{4.7}$$

If we subtract line (4.7) from (4.6), on the left side, we'll get $\vec{x} - \vec{x} = \vec{0}_n$. Combining like terms on the right, we get

$$\vec{0}_n = (a_1 - c_1)\vec{b}_1 + (a_2 - c_2)\vec{b}_2 + \cdots + (a_k - c_k)\vec{b}_k.$$

But $\mathcal{B}$ is a basis, so it is linearly independent, meaning that the only solution of this homogeneous equation is the trivial solution,

$$\begin{array}{rcl} a_1 - c_1 & = & 0 \\ a_2 - c_2 & = & 0 \\ \vdots & \vdots & \vdots \\ a_k - c_k & = & 0 \end{array}.$$

That is, the coefficients are actually the same, $a_i = c_i$ for each $i = 1, \ldots, k$. So, once an ordered basis is specified for some subspace of R^n (including R^n itself), any element of the subspace can be uniquely identified with the coefficients for its representation as a linear combination of the basis elements.

Remark 4.3.1. *In Example 1.1.1, we saw that* $\text{Span}\{\vec{e}_1\}$*, where* $\vec{e}_1 = \langle 1, 0\rangle$ *in* R^2*, could be equated with the horizontal axis. It is easy to show that* $\mathcal{B} = \{\vec{e}_1\}$ *is a basis for this subspace of* R^2*. Early in Section 4.2, we mentioned that this subset (which is a subspace) somehow has a similar structure to* R^1*. Note that given any vector in* $\text{Span}\{\vec{e}_1\}$*, we can associate it in a unique way with its coefficient. For example, we can equate the real number* 2 *with the vector* $\langle 2, 0\rangle$*, the real number* π *with the vector* $\langle \pi, 0\rangle$*, or more generally, the real number* c *with the vector* $\langle c, 0\rangle$*. This is the sense in which* $\mathcal{B} = \{\vec{e}_1\}$ *has the same structure as* R^1*. As long as we know that our context is the basis* $\mathcal{B} = \{\vec{e}_1\}$*, we can represent each element in the subspace with a single element of* R^1*.*

Since the coefficients are uniquely determined once we've specified a basis for a subspace of R^n, we can work with the set of coefficients. In fact, these coefficients can be used to create a new set of k-tuples that define the elements of our subspace. We call these **coordinate vectors**.

Definition 4.3.2. *Let S be a subspace of R^n and $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_k\}$ be an ordered basis of S. For each element $\vec{x}$ in S, the* **coordinate vector for $\vec{x}$ relative to the basis $\mathcal{B}$** *is denoted $[\vec{x}]_{\mathcal{B}}$ and is defined to be*

$$[\vec{x}]_{\mathcal{B}} = \langle c_1, c_2, \ldots, c_k \rangle,$$

where the entries are the coefficients of the representation of $\vec{x}$ as a linear combination of the basis elements. That is, the c's are the coefficients in the equation

$$\vec{x} = c_1\vec{b}_1 + c_2\vec{b}_2 + \cdots + c_k\vec{b}_k.$$

Example 4.3.7. *For the subspace $X = \text{Span}\{\vec{e}_1\}$ of R^2 with basis $\mathcal{B} = \{\vec{e}_1\}$, the coordinate vectors are real numbers. For example,*

$$[\langle 2, 0\rangle]_{\mathcal{B}} = 2, \quad [\langle \pi, 0\rangle]_{\mathcal{B}} = \pi, \quad \textit{and} \quad [\langle c, 0\rangle]_{\mathcal{B}} = c.$$

We can write elements of R^1 using brackets, i.e., $\langle c\rangle$, though it is not customary to do so.

Example 4.3.8. *Consider the basis $\mathcal{B} = \{\langle 2, 1\rangle, \langle 1, 2\rangle\}F$ of R^2 (we showed in Exercise 4.3.3 that this is a basis).*

1. *Find $[\vec{x}]_{\mathcal{B}}$ if $\vec{x} = \langle 2, 1\rangle$*
2. *Find $[\vec{x}]_{\mathcal{B}}$ if $\vec{x} = \langle 1, 2\rangle$*
3. *Find $[\vec{x}]_{\mathcal{B}}$ if $\vec{x} = \langle 0, 0\rangle$*
4. *Find $[\vec{x}]_{\mathcal{B}}$ if $\vec{x} = \langle 1, 0\rangle$*
5. *Find $\vec{x}$ if $[\vec{x}]_{\mathcal{B}} = \langle 3, 1\rangle$*

To find each coordinate vector, we can translate the problem into an equation which we can solve. Some of these can be done with little effort. For 1., note that we seek

$$[\langle 2, 1\rangle]_{\mathcal{B}} = \langle c_1, c_2\rangle$$

where

$$\langle 2, 1\rangle = c_1\langle 2, 1\rangle + c_2\langle 1, 2\rangle.$$

By observation, we have $c_1 = 1$ and $c_2 = 0$. So

$$[\langle 2, 1\rangle]_{\mathcal{B}} = \langle 1, 0\rangle.$$

A similar argument can be applied to 2. to arrive at

$$[\langle 1,2\rangle]_{\mathcal{B}} = \langle 0,1\rangle.$$

For 3., we can consider the equation

$$\langle 0,0\rangle = c_1\langle 2,1\rangle + c_2\langle 1,2\rangle.$$

Since the basis elements are linearly independent, we must have

$$[\langle 0,0\rangle]_{\mathcal{B}} = \langle 0,0\rangle.$$

Exercise 4. will require a bit more effort, but we can start with the equation

$$\langle 1,0\rangle = c_1\langle 2,1\rangle + c_2\langle 1,2\rangle.$$

We can rephrase the problem in terms of a matrix-vector equation $B\vec{c} = \vec{x}$*, and our coordinate vector* $[\vec{x}]_{\mathcal{B}}$ *will be the solution* $\vec{c}$*. The matrix* B *will have columns* $\langle 2,1\rangle$ *and* $\langle 1,2\rangle$ *(in this order). Setting up an augmented matrix and using row reduction,*

$$[B\,|\,\vec{x}] = \left[\begin{array}{cc|c} 2 & 1 & 1 \\ 1 & 2 & 0 \end{array}\right] \quad \xrightarrow{rref} \quad \left[\begin{array}{cc|c} 1 & 0 & 2/3 \\ 0 & 1 & -1/3 \end{array}\right].$$

It follows that

$$[\langle 1,0\rangle]_{\mathcal{B}} = \left\langle \frac{2}{3}, -\frac{1}{3} \right\rangle.$$

We can readily verify that this is correct,

$$\frac{2}{3}\langle 2,1\rangle - \frac{1}{3}\langle 1,2\rangle = \left\langle \frac{4}{3} - \frac{1}{3}, \frac{2}{3} - \frac{2}{3} \right\rangle = \langle 1,0\rangle.$$

Exercise 5. is a bit different. Here, we are given the coefficients, $c_1 = 3$ *and* $c_2 = 1$*, we simply need to find the linear combination of our basis vectors.*

$$\vec{x} = 3\langle 2,1\rangle + 1\langle 1,2\rangle = \langle 3(2)+1, 3(1)+2\rangle = \langle 7,5\rangle.$$

Exercise 5. in Example 4.3.8 above illustrates that a vector $\vec{x}$ in some subspace S of R^n can be realized as a matrix-vector product, where the vector is the coordinate vector, and the matrix has the basis elements for columns.

That is, if we form a matrix B whose columns are the basis elements (in the order given), then

$$\vec{x} = B[\vec{x}]_{\mathcal{B}}.$$

The basis in Example 4.3.8 would give us the matrix

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

So we could determine the vector $\vec{x}$ in part 5 of Example 4.3.8 by evaluating the matrix-vector product

$$\vec{x} = B[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \langle 3, 1 \rangle = \langle 7, 5 \rangle.$$

If B is a square matrix, we can also find coordinate vectors using the relationship

$$[\vec{x}]_{\mathcal{B}} = B^{-1}\vec{x}.$$

(Can you explain how we would know that such a square matrix B would be invertible?) In general, however, the matrix formed from the basis elements need not be square. If S is a subspace of R^n with basis $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_k\}$, then the matrix B defined by

$$\mathrm{Col}_i(B) = \vec{b}_i$$

will have size $n \times k$.

Exercise 4.3.4. *Consider the basis $\mathcal{B} = \{\langle 1, 1 \rangle, \langle 1, -1 \rangle\}$ of R^2.*

1. *Find $[\vec{x}]_{\mathcal{B}}$ if $\vec{x} = \langle 1, 1 \rangle$*

2. *Find $[\vec{x}]_{\mathcal{B}}$ if $\vec{x} = \langle 1, -1 \rangle$*

3. *Find $[\vec{x}]_{\mathcal{B}}$ if $\vec{x} = \langle 0, 0 \rangle$*

4. *Find $[\vec{x}]_{\mathcal{B}}$ if $\vec{x} = \langle 2, 3 \rangle$*

5. *Find $\vec{x}$ if $[\vec{x}]_{\mathcal{B}} = \langle -1, 4 \rangle$*

Example 4.3.9. *Consider the vectors* $\vec{e}_1 = \langle 1, 0, 0, 0\rangle$ *and* $\vec{e}_3 = \langle 0, 0, 1, 0\rangle$ *in* R^4*. Let* $\mathcal{B} = \{\vec{e}_1, \vec{e}_3\}$ *be an ordered basis for the subspace* $P = \text{Span}\{\mathcal{B}\}$*. Describe the coordinate vectors for elements of* P*.*

If $\vec{x}$ *is in* P*, then* $\vec{x} = \langle x_1, 0, x_3, 0\rangle$ *for some real numbers* x_1 *and* x_3*. The coordinate vector for* $\vec{x}$ *relative to* $\mathcal{B}$ *will be*

$$[\vec{x}]_{\mathcal{B}} = \langle x_1, x_3\rangle.$$

We note that the coordinate vectors will be vectors in R^2*. In fact, given any vector, say* $\vec{y} = \langle y_1, y_2\rangle$ *in* R^2*, the vector* $\vec{x} = \langle y_1, 0, y_2, 0\rangle$ *will be an element of* P *(since it can be written as* $y_1\vec{e}_1 + y_2\vec{e}_3$ *in* R^4*). So the set of all coordinate vectors of* P *relative to the basis* $\mathcal{B}$ *is actually all of* R^2*.*

Exercise 4.3.5. *For the subspace* P *in Example 4.3.9, construct the matrix* B *whose columns are the basis elements in the order given. For each coordinate vector* $[\vec{x}]_{\mathcal{B}}$ *in* R^2*, find the element* $\vec{x}$ *in* P *by using the matrix-vector product* $\vec{x} = B[\vec{x}]_{\mathcal{B}}$*.*

1. $[\vec{x}]_{\mathcal{B}} = \langle 1, 1\rangle$
2. $[\vec{x}]_{\mathcal{B}} = \langle -3, 5\rangle$
3. $[\vec{x}]_{\mathcal{B}} = \langle x_1, x_2\rangle$

Remark 4.3.2. *The observation in Example 4.3.9, that the collection of coordinate vectors for the subspace* P *of* R^4 *is the space* R^2*, has a name. We say that* P *is* **isomorphic** *to* R^2*. Similarly, from Example 4.3.7, we can say that* X *is isomorphic to* R^1*.*

Exercise 4.3.6. *Consider the vectors* $\vec{b}_1$*,* $\vec{b}_2$*, and* $\vec{b}_3$ *in* R^5 *given by*

$$\vec{b}_1 = \langle 1, 0, 0, 0, 0\rangle, \quad \vec{b}_2 = \langle 1, 1, 0, 0, 0\rangle, \quad \text{and} \quad \vec{b}_3 = \langle 1, 1, 1, 0, 0\rangle.$$

Let $\mathcal{B} = \left\{\vec{b}_1, \vec{b}_2, \vec{b}_3\right\}$ *be the ordered basis for the subspace* $S = \text{Span}(\mathcal{B})$ *of* R^5*. Verify that the coordinate vectors,* $[\vec{b}_i]_{\mathcal{B}}$*, of the basis elements are the standard unit vectors in* R^3*. That is, show that*

$$[\vec{b}_1]_{\mathcal{B}} = \langle 1, 0, 0\rangle, \quad [\vec{b}_2]_{\mathcal{B}} = \langle 0, 1, 0\rangle, \quad \text{and} \quad [\vec{b}_3]_{\mathcal{B}} = \langle 0, 0, 1\rangle.$$

Exercise 4.3.7. *Suppose $n \geq 2$ and $1 \leq k \leq n$. Let $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_k\}$ be an ordered basis of the subspace $S = \text{Span}(\mathcal{B})$ of R^n. Explain why $[\vec{b}_i]_{\mathcal{B}} = \vec{e}_i$ for each $i = 1, \ldots, k$. That is, explain why the coordinate vectors for the basis elements are the standard unit vectors in R^k.*

Hint: don't worry about a bunch of computations, just consider the equation

$$\vec{b}_i = c_1\vec{b}_1 + \cdots + c_i\vec{b}_i + \cdots + c_k\vec{b}_k.$$

4.3.2 Dimension

We are ready to define the dimension of R^n or more generally the dimension of a subspace of R^n. Fortunately, the definition of dimension in this context will align with our intuition—i.e., that R^2 should be two dimensional, R^3 should be three dimensional, and so forth. But we need a rigorous and unambiguous criterion upon which to base the definition of dimension. This will come from bases.

Theorem 4.3.1. *Let $n \geq 2$ and $1 \leq k \leq n$. Suppose S is a subspace of R^n and $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_k\}$ is a basis of S. Then every basis of S consists of exactly k vectors.*

To prove Theorem 4.3.1, we first establish the following lemma.

Lemma 4.3.1. *Suppose S is a subspace of R^n and $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_k\}$ is an ordered basis of S. If $T = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ is any set of m vectors in S where $m > k$, then T is linearly dependent.*

Proof. We will use coordinate vectors to prove Lemma 4.3.1. First, note that since $\mathcal{B}$ consists of k vectors, coordinate vectors with respect to $\mathcal{B}$ will be vectors in R^k. Let A be the $k \times m$ matrix whose columns are the coordinate vectors of the elements of T with respect to the basis $\mathcal{B}$,

$$\text{Col}_i(A) = [\vec{v}_i]_{\mathcal{B}}, \quad i = 1, \ldots, m.$$

Now, A has m columns, each of which is a vector in R^k, and $m > k$. By Theorem 4.1.2, the columns of A are linearly dependent. So applying Theorem 4.1.1, the homogeneous equation $A\vec{x} = \vec{0}_k$ has a nontrivial solution, $\vec{x} = \langle x_1, x_2, \ldots, x_m \rangle$. Hence we have a linear dependence relation,

$$x_1[\vec{v}_1]_{\mathcal{B}} + x_2[\vec{v}_2]_{\mathcal{B}} + \cdots + x_m[\vec{v}_m]_{\mathcal{B}} = \vec{0}_k \tag{4.8}$$

with at least one $x_i \neq 0$, for the coordinate vectors in R^k. We can use equation (4.8) to obtain a linear dependence relation on the vectors in T. To this end, we create the $n \times k$ matrix B whose columns are the basis elements in $\mathcal{B}$,

$$\mathrm{Col}_i(B) = \vec{b}_i, \quad i = 1, \ldots, k.$$

By the definition of coordinate vectors,

$$\vec{v}_i = B[\vec{v}_i]_{\mathcal{B}} \quad \text{for each} \quad \vec{v}_i \in T.$$

We multiply both sides of equation (4.8) by the matrix B and make use of the distributive property to obtain

$$\begin{array}{ccccccccc} B\Big(x_1[\vec{v}_1]_{\mathcal{B}} & + & x_2[\vec{v}_2]_{\mathcal{B}} & + & \cdots & + & x_m[\vec{v}_m]_{\mathcal{B}}\Big) & = & B\vec{0}_k \\ x_1B[\vec{v}_1]_{\mathcal{B}} & + & x_2B[\vec{v}_2]_{\mathcal{B}} & + & \cdots & + & x_mB[\vec{v}_m]_{\mathcal{B}} & = & \vec{0}_n \\ x_1\vec{v}_1 & + & x_2\vec{v}_2 & + & \cdots & + & x_m\vec{v}_m & = & \vec{0}_n \end{array} . \tag{4.9}$$

Equation (4.9) is a linear dependence relation, and we conclude that T is linearly dependent. □

We can now prove prove Theorem 4.3.1.

Proof. (of Theorem 4.3.1) Suppose a subspace S of R^n has two bases $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_k\}$ and $\mathcal{C} = \{\vec{c}_1, \ldots, \vec{c}_\ell\}$ consisting of k and ℓ vectors, respectively. By Lemma 4.3.1, as $\mathcal{C}$ must be linearly independent, we have $\ell \leq k$. Similarly, the linear independence of $\mathcal{B}$ requires $k \leq \ell$. Hence it must be that $k = \ell$, and we conclude that every basis of a subspace S of R^n must contain the same number of elements. □

We are ready to define dimension.

Definition 4.3.3. *Let S be a subspace of R^n. If $S = \{\vec{0}_n\}$, then the dimension of S, written $\dim(S)$ is equal to zero. If S contains more than the zero vector, then the dimension of S, $\dim(S) = k$, where k is the number of elements in any basis of S.*

The subset of R^n containing only the zero vector is a subspace of R^n (see Exercise 2). This subspace does not have a basis, so we assign its dimension to be zero. Otherwise, we have established that every basis for a given subspace must contain the same number of basis vectors. So the dimension is well defined. As we expected, the dimension of R^n is n.

Example 4.3.10. *Show that for any $n \geq 2$, $\dim(R^n) = n$.*

For any $n \geq 2$ we can take the standard basis for R^n consisting of the n vectors $\vec{e}_i$, having a 1 in the i^{th} entry and zero in all other entries. Since this set consists of n vectors, $\dim(R^n) = n$.

Example 4.3.11. *Find the dimension of the null space of the matrix*

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

from Example 4.3.6.

In Example 4.3.6, we found the basis $\{\langle -2, 1, 1, 0\rangle, \langle 1, -2, 0, 1\rangle\}$ for $\mathcal{N}(A)$. Since this basis contains two vectors, we conclude that

$$\dim(\mathcal{N}(A)) = 2.$$

Exercise 4.3.8. *Find the dimension of the null space of the matrix*

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

Exercise 4.3.9. *Find the dimension of the null space of the matrix*

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

4.3.3 Basis as a Subset of a Spanning Set

We are trying to make our way to a profound result in linear algebra called the *Fundamental Theorem of Linear Algebra.* (If they call it "fundamental," it must be important, right?) This theorem is actually a collection of related results that in part connects the dimensions of the fundamental subspaces of a matrix. Right now, we have a reliable process for characterizing the null space of a given matrix, either as the space $\{\vec{0}_n\}$ or having some basis that we can find. However, we don't yet have a method to find a basis for the row or column spaces of a matrix. For example, given an $m \times n$ matrix, A, we can take the set of row vectors as a spanning set for $\mathcal{RS}(A)$ (similarly the column vectors are a spanning set for $\mathcal{CS}(A)$). But such a set is not necessarily a basis, since the row vectors need not be linearly independent. We can ask, if we're given a spanning set that is not linearly independent, is there a reliable way to cull the set so that we (1) span the same subspace, and (2) obtain a linearly independent spanning set? We begin by proving the following result that says that we can cull a linearly dependent set (in specific ways) without changing the subspace it spans.

Lemma 4.3.2. *Let $T = \{\vec{u}_1, \ldots, \vec{u}_k\}$ be a set of k vectors in R^n, with $k \geq 2$, and let $S = \text{Span}(T)$. If one of the vectors, say $\vec{u}_i$ in T, is a linear combination of the other vectors in T, then the set obtained from T by removing $\vec{u}_i$ spans S.*

We might recognize the general idea stated in this Lemma from Example 4.3.1. In that example, we showed that the span remained the same when we removed a vector from a linearly dependent set. Here, we prove Lemma 4.3.2.

Proof. Suppose one of the vectors in T is a linear combination of the other vectors in T,

$$\vec{u}_i = c_1\vec{u}_1 + \cdots + c_{i-1}\vec{u}_{i-1} + c_{i+1}\vec{u}_{i+1} + \cdots + c_k\vec{u}_k, \tag{4.10}$$

and let $\hat{T}$ be the subset of T obtained by removing $\vec{u}_i$. Since $\hat{T} \subset T$, any vector $\vec{x}$ in $\text{Span}(\hat{T})$ will be a linear combination of vectors in T. Hence $\text{Span}(\hat{T}) \subset \text{Span}(T)$. So suppose $\vec{x}$ is in $\text{Span}(T)$. Then we can write $\vec{x}$ as a linear combination of the vectors in T,

$$\vec{x} = a_1\vec{u}_1 + \cdots + a_{i-1}\vec{u}_{i-1} + a_i\vec{u}_i + a_{i+1}\vec{u}_{i+1} + \cdots + a_k\vec{u}_k. \tag{4.11}$$

Now, we can replace $\vec{u}_i$ in equation (4.11) with the right hand side of equation (4.10) and collect terms to obtain

$$\begin{aligned} \vec{x} = (a_1 + a_ic_1)\vec{u}_1 \quad &+ \quad \cdots + (a_{i-1} + a_ic_{i-1})\vec{u}_{i-1} + \\ &+ \quad (a_{i+1} + a_ic_{i+1})\vec{u}_{i+1} + \cdots + (a_k + a_ic_k)\vec{u}_k. \end{aligned} \tag{4.12}$$

Hence $\vec{x}$ is a linear combination of the vectors in $\hat{T}$ so that $\text{Span}(T) \subset \text{Span}(\hat{T})$. It follows that $\text{Span}(\hat{T}) = \text{Span}(T)$, and the proof is complete. □

Now, we show that a basis can be obtained from a spanning set.

Lemma 4.3.3. *Let $T = \{\vec{u}_1, \ldots, \vec{u}_k\}$ be a set of vectors in R^n and $S = \mathrm{Span}(T)$. If T contains at least one nonzero vector, then there exists a subset of T that is a basis for S.*

Proof. If T contains at least one nonzero vector, then S contains nonzero vectors and hence has a basis. If T is linearly independent, then T is a basis, and we are done. Otherwise, we can apply Lemma 4.3.2, repeating as necessary, to remove vectors until we are left with a linearly independent set. This will be a basis for S. □

The Lemmas 4.3.2 and 4.3.3 tell us that we can obtain a basis from a spanning set that is not linearly independent. In particular, they say that we can remove vectors that are known to be linear combinations of the other vectors. If our goal is to find a basis, say for the column and row spaces of a matrix, this is encouraging because we have spanning sets. Unfortunately, these Lemmas don't give us any practical advice on how to do this. With two or three vectors, we might be able to find linear dependence relationships by just looking closely at the entries, but what if we're dealing with a 20×30 matrix? It's too much to ask that we can just *see* linear dependence relations. Well, maybe it's not always too much to ask. If a matrix has a particularly nice structure, as Example 4.3.12 illustrates, it may be possible to identify linear dependence relationships by simple observation.

Example 4.3.12. *For the rref matrix A in equation (4.13), find a basis for $\mathcal{CS}(A)$.*

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & -3 & 0 & 1 \\ 0 & 0 & 0 & 1 & 5 & 0 & -4 \\ 0 & 0 & 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{4.13}$$

The matrix A has four pivot columns, the first, third, fourth and sixth. The set of pivot columns, $\{\mathrm{Col}_1(A), \mathrm{Col}_3(A), \mathrm{Col}_4(A), \mathrm{Col}_6(A)\}$ is obviously linearly independent—since each has a 1 in a position where all the others have zero. Since these pivot columns are standard unit vectors, we can immediately see how the non-pivot columns depend on the pivot columns. By *observation, we see that*

$$\begin{aligned}
\mathrm{Col}_2(A) &= -\mathrm{Col}_1(A), &(4.14)\\
\mathrm{Col}_5(A) &= 2\,\mathrm{Col}_1(A) - 3\,\mathrm{Col}_3(A) + 5\,\mathrm{Col}_4(A), \quad \textit{and} &(4.15)\\
\mathrm{Col}_7(A) &= \mathrm{Col}_3(A) - 4\,\mathrm{Col}_4(A) + 7\,\mathrm{Col}_6(A). &(4.16)
\end{aligned}$$

Note that the coefficients of the pivot columns in equations (4.14)–(4.16) are simply the entries in the corresponding non-pivot columns—e.g., the coefficients in equation (4.15) are the numbers 2*,* -3 *and* 5 *from that* 5^{th} *column. We start with a spanning set that contains all of the columns,*

$$\mathcal{CS}(A) = \mathrm{Span}\{\mathrm{Col}_1(A), \mathrm{Col}_2(A), \mathrm{Col}_3(A), \mathrm{Col}_4(A), \mathrm{Col}_5(A), \mathrm{Col}_6(A), \mathrm{Col}_7(A)\},$$

and noting the linear dependence demonstrated in equations (4.14)–(4.16), we apply Lemma 4.11 three times to remove $\mathrm{Col}_2(A)$*,* $\mathrm{Col}_5(A)$*, and* $\mathrm{Col}_7(A)$*. We are left with a basis*

$$\begin{aligned}
\textit{Basis for } \mathcal{CS}(A) &= \{\mathrm{Col}_1(A), \mathrm{Col}_3(A), \mathrm{Col}_4(A), \mathrm{Col}_6(A)\}\\
&= \{\langle 1,0,0,0,0\rangle, \langle 0,1,0,0,0\rangle, \langle 0,0,1,0,0\rangle, \langle 0,0,0,0,1\rangle\}.
\end{aligned}$$

The critical conclusion in Example 4.3.12 is that **the set of pivot columns constitute a basis for** $\mathcal{CS}(A)$. Of course, the pivot columns in any rref are standard unit vectors, so the procedure we used in this example applies to any rref. We can even make the following generalization:

Remark 4.3.3. *If* A *is a nonzero matrix that is an rref, then the pivot columns of* A *form a basis for* $\mathcal{CS}(A)$*.*

Remark 4.3.3 smells like a theorem—it could be stated as a theorem—and it's easy to argue that it is true. But it is very restrictive. We want to be able to find a basis for the column space of a matrix without confining ourselves to rrefs. In fact, given any set of vectors in R^m, we can always use them to define a matrix (using the vectors as columns). So, if we have a method for finding the basis of a column space, we will have a method for finding a basis of **any** set of vectors. Remarkably (pun intended), the conclusion of Remark 4.3.3 is true even if we drop the condition that A is an rref. This will be the main result of the next section.

Exercise 4.3.10. *Consider the matrix* H*.*

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

1. *Classify each column as a pivot column or a non-pivot column.*

2. *Express each non-pivot column as a linear combination of one or more pivot columns.*

3. *Identify a basis for $\mathcal{CS}(H)$.*

4.4 Bases for the Column and Row Spaces of a Matrix

4.4.1 Basis for a Column Space

We begin with the main theorem of this section.

Theorem 4.4.1. *Let A be an $m \times n$ matrix that is not the zero matrix. The pivot columns of A form a basis for $\mathcal{CS}(A)$.*

Before we prove Theorem 4.4.1, let's look back at how we solve matrix-vector equations—or systems of linear equations—using the rref, and the relationship between basic and free variables.

Example 4.4.1. *Consider the matrix-vector equation $A\vec{x} = \vec{y}$ where*

$$A = \begin{bmatrix} 1 & -2 & 0 & 1 & 8 \\ 2 & -4 & 3 & 0 & -6 \\ -1 & 2 & 1 & 2 & 3 \\ 3 & -6 & 4 & 0 & -7 \end{bmatrix}, \quad \textit{and} \quad \vec{y} = \langle 2, -1, 11, -2 \rangle.$$

We set up the augmented matrix $\left[A \,|\, \vec{y}\right]$ and reduced the result to an rref having the form $\left[\text{rref}(A) \,|\, \vec{z}\right]$

$$\left[\begin{array}{rrrrr|r} 1 & -2 & 0 & 1 & 8 & 2 \\ 2 & -4 & 3 & 0 & -6 & -1 \\ -1 & 2 & 1 & 2 & 3 & 11 \\ 3 & -6 & 4 & 0 & -7 & -2 \end{array}\right] \xrightarrow{rref} \left[\begin{array}{rrrrr|r} 1 & -2 & 0 & 0 & 3 & -2 \\ 0 & 0 & 1 & 0 & -4 & 1 \\ 0 & 0 & 0 & 1 & 5 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

From the rref, we can identify the basic and free variables, and deduce an equation of the form

$$x_1 \operatorname{Col}_1(A) + x_2 \operatorname{Col}_2(A) + x_3 \operatorname{Col}_3(A) + x_4 \operatorname{Col}_4(A) + x_5 \operatorname{Col}_5(A) = \vec{y}. \quad (4.17)$$

For this particular example, we see that x_1, x_3 *and* x_4 *are basic variables, and* x_2 *and* x_5 *are free. Specifically, a parametric representation of the solution is*

$$\begin{array}{rcl} x_1 & = & -2+2s-3t \\ x_2 & = & s \\ x_3 & = & 1+4t \\ x_4 & = & 4-5t \\ x_5 & = & t \end{array} \quad , \quad -\infty < s,t < \infty.$$

Substituting these into equation (4.17), we express $\vec{y}$ *as a linear combination of the columns of* A

$$(-2+2s-3t)\,\mathrm{Col}_1(A) + s\,\mathrm{Col}_2(A) + (1+4t)\,\mathrm{Col}_3(A) + (4-5t)\,\mathrm{Col}_4(A) + t\,\mathrm{Col}_5(A) = \vec{y}. \tag{4.18}$$

We note that this equation is true for all choices of the parameters s *and* t*. In particular, taking* $s = t = 0$*, we see that we can write* $\vec{y}$ *as*

$$-2\,\mathrm{Col}_1(A) + 1\,\mathrm{Col}_3(A) + 4\,\mathrm{Col}_4(A) = \vec{y}. \tag{4.19}$$

The critical feature of equation (4.19) is that we have expressed the vector $\vec{y}$ as a linear combination of the pivot columns alone. That is, since we can set all free variables to zero (which is legitimate because they are free), we can exclude the non-pivot columns and still express our solution.

Let's recall that for an $m \times n$ matrix A, one interpretation of the column space, $\mathcal{CS}(A)$, is the set of all $\vec{y}$ in R^m such that $A\vec{x} = \vec{y}$ is consistent. What we see in this example can be generalized to argue Theorem 4.4.1.

Proof. (Of Theorem 4.4.1.) Let A be an $m \times n$ matrix that is not the zero matrix. Then $\mathcal{CS}(A)$ contains nonzero vectors and hence has a basis. Let $\mathcal{B}_{pcol}$ be the set of pivot columns of A. Since A is not the zero matrix, $\mathcal{B}_{pcol}$ is a non-empty subset of the set of column vectors of A, and because $\mathcal{B}_{pcol}$ contains only the pivot column vectors of A, $\mathcal{B}_{pcol}$ is linearly independent. Since $\mathcal{B}_{pcol}$ is a subset of the set of column vectors of A, $\mathrm{Span}\{\mathcal{B}_{pcol}\} \subseteq \mathcal{CS}(A)$. Now, let $\vec{y}$ be any element of $\mathcal{CS}(A)$. Then the equation $A\vec{x} = \vec{y}$ is consistent. Let $\vec{x}_0$ be the solution to this equation for which all free variables (if any) are set to zero. Since free variables are the coefficients of the non-pivot columns of A, $A\vec{x}_0$ is a linear combination of the pivot columns of A. But $A\vec{x}_0 = \vec{y}$ so that $\vec{y} \in \mathrm{Span}\{\mathcal{B}_{pcol}\}$. Hence $\mathcal{CS}(A) \subseteq \mathrm{Span}\{\mathcal{B}_{pcol}\}$. It follows that $\mathcal{CS}(A) = \mathrm{Span}\{\mathcal{B}_{pcol}\}$, and hence $\mathcal{B}_{pcol}$ is a basis for $\mathcal{CS}(A)$. □

Example 4.4.2. *Find a basis for $\mathcal{CS}(A)$ for the matrix*

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

Theorem 4.4.1 says that the set of pivot columns of A is a basis for $\mathcal{CS}(A)$. To identify the pivot columns, we perform row reduction.

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

From $\mathrm{rref}(A)$, *we see that the first two columns of A are pivot columns. So we can take as a basis the set consisting of the first two column vectors of A. Calling the basis $\mathcal{B}$, we have*

$$\mathcal{B} = \{\langle 0, 4, -2, 5\rangle, \langle 3, 7, -5, -4\rangle\}.$$

Remark 4.4.1. *To find a basis for $\mathcal{CS}(A)$, we use row reduction in order to identify which columns of A are pivot columns. We need to remember that the actual vectors that will be our basis elements are column vectors of A, not* $\mathrm{rref}(A)$*! For example, in Example 4.4.2, it would be incorrect to take the first two columns of* $\mathrm{rref}(A)$ *as a basis for $\mathcal{CS}(A)$. This is actually quite obvious in this example. Every linear combination of the first two columns of* $\mathrm{rref}(A)$ *will necessarily have third and fourth entry zero. Even the columns of A itself have nonzero third and fourth entries!*

Exercise 4.4.1. *Find a basis for the column space of each matrix.*

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

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

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

4.4.2 Basis for a Row Space

Given that the columns of A^T are the rows of A, one method for obtaining a basis for $\mathcal{RS}(A)$, given any nonzero matrix A, is to find a basis for $\mathcal{CS}(A^T)$. This is a perfectly legitimate approach. The only possible drawback to using this in practice is that it requires another row reduction procedure. (Of course, this is not such an imposition when using technology as most matrix manipulation software will have a transpose command in addition to row reduction functions.) It is possible to glean a basis for the row space of a matrix A from $\text{rref}(A)$ without performing row reduction on A^T. And this follows from the nature of row operations and their effect on a spanning set. In particular, elementary row operations preserve the row space.

Theorem 4.4.2. *If A and B are row equivalent matrices, then $\mathcal{RS}(A) = \mathcal{RS}(B)$.*

We'll recall that two matrices are row equivalent if one can be obtained from the other by performing some sequence of elementary row operations. In particular, A and $\text{rref}(A)$ are row equivalent. To prove Theorem 4.4.2, we consider each of the three elementary row operations and their effects on a span. The first is rather obvious.

Lemma 4.4.1. *Let A be an $m \times n$ matrix, and let B be the $m \times n$ matrix obtained from A by performing a row swap, $R_i \leftrightarrow R_j$ for some i and j in $\{1, \ldots, m\}$. Then $\mathcal{RS}(A) = \mathcal{RS}(B)$.*

Proof. If B is obtained from A by performing a row swap, then the sets of row vectors,

$$\{\text{Row}_1(A), \ldots, \text{Row}_m(A)\} = \{\text{Row}_1(B), \ldots, \text{Row}_m(B)\}.$$

While the labels may be different, the sets contain the same vectors. Hence

$$\text{Span}\{\text{Row}_1(A), \ldots, \text{Row}_m(A)\} = \text{Span}\{\text{Row}_1(B), \ldots, \text{Row}_m(B)\}.$$

□

Lemma 4.4.2. *Let A be an $m \times n$ matrix, and let B be the $m \times n$ matrix obtained from A by performing a row scaling, $kR_i \to R_i$ for some i in $\{1, \dots, m\}$. Then $\mathcal{RS}(A) = \mathcal{RS}(B)$.*

Proof. Let A and B be $m \times n$ matrices and suppose B is obtained from A by scaling the i^{th} row by the (necessarily nonzero) number k. Let $\vec{x} \in \mathcal{RS}(A)$, then $\vec{x}$ can be written as a linear combination of the row vectors of A. But note then that

$$\begin{aligned} \vec{x} &= c_1 \operatorname{Row}_1(A) + \dots + c_i \operatorname{Row}_i(A) + \dots + c_m \operatorname{Row}_m(A) \\ &= c_1 \operatorname{Row}_1(A) + \dots + \left(\tfrac{c_i}{k}\right) k \operatorname{Row}_i(A) + \dots + c_m \operatorname{Row}_m(A) \\ &= c_1 \operatorname{Row}_1(B) + \dots + \left(\tfrac{c_i}{k}\right) \operatorname{Row}_i(B) + \dots + c_m \operatorname{Row}_m(B), \end{aligned} \tag{4.20}$$

showing that $\vec{x}$ can be written as a linear combination of the row vectors of B, i.e., $\vec{x} \in \mathcal{RS}(B)$. Equation (4.20) similarly shows that $\mathcal{RS}(A) \subseteq \mathcal{RS}(B)$, and we can conclude that $\mathcal{RS}(A) = \mathcal{RS}(B)$. □

Lemma 4.4.3. *Let A be an $m \times n$ matrix, and let B be the $m \times n$ matrix obtained from A by performing a row replacement, $kR_i + R_j \to R_j$ for some i and j in $\{1, \dots, m\}$. Then $\mathcal{RS}(A) = \mathcal{RS}(B)$.*

Proof. Let A be an $m \times n$ matrix, and let B be the $m \times n$ matrix obtained from A by replacing the j^{th} row of A with $kR_i + R_j$ for some row R_i and scalar k. Let $\vec{x} \in \mathcal{RS}(A)$. Then we can write $\vec{x}$ as a linear combination of the rows of A. For ease of notation, let $\vec{R}_{iA} = \operatorname{Row}_i(A)$ and $\vec{R}_{iB} = \operatorname{Row}_i(B)$. Note that

$$\begin{aligned} \vec{x} &= c_1\vec{R}_{1A} + \dots + c_i\vec{R}_{iA} + \dots + c_j\vec{R}_{jA} + c_m\vec{R}_{mA} \\ &= c_1\vec{R}_{1A} + \dots + c_i\vec{R}_{iA} + \dots + c_j\vec{R}_{jA} + c_m\vec{R}_{mA} + kc_j\vec{R}_{iA} - kc_j\vec{R}_{iA} \\ &= c_1\vec{R}_{1A} + \dots + (c_i - kc_j)\vec{R}_{iA} + \dots + c_j\left(k\vec{R}_{iA} + \vec{R}_{jA}\right) + c_m\vec{R}_{mA} \\ &= c_1\vec{R}_{1B} + \dots + (c_i - kc_j)\vec{R}_{iB} + \dots + c_j\vec{R}_{jB} + c_m\vec{R}_{mB}, \end{aligned} \tag{4.21}$$

showing that $\vec{x}$ can be written as a linear combination of the row vectors of B, i.e., $\vec{x} \in \mathcal{RS}(B)$. Equation (4.21) similarly shows that $\mathcal{RS}(A) \subseteq \mathcal{RS}(B)$, and we can conclude that $\mathcal{RS}(A) = \mathcal{RS}(B)$. □

The proof of Theorem 4.4.2 immediately follows.

Proof. (of Theorem 4.4.2) Suppose A and B are $m \times n$ matrices that are row equivalent. Then A can be transformed into B by some finite sequence of elementary row operations. Apply Lemmas 4.4.1, 4.4.2, and 4.4.3, repeating as necessary. □

The main reason that Theorem 4.4.2 is useful is that the nonzero rows of any echelon form are necessarily linearly independent—due to the placement of ones and zeros. Hence, given a matrix A:

The nonzero rows of $\text{rref}(A)$ form a basis for $\mathcal{RS}(A)$.

Example 4.4.3. *Find a basis for $\mathcal{RS}(A)$ for the matrix*

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

To find a basis for the column space in Example 4.4.2, we found $\text{rref}(A)$

$$\text{rref}(A) = \begin{bmatrix} 1 & 0 & 2/3 \\ 0 & 1 & 1/3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

By Theorem 4.4.2, we can take the nonzero rows of this echelon form as our basis vectors. Calling the basis $\mathcal{B}$, a solution is

$$\mathcal{B} = \{\langle 1, 0, 2/3 \rangle, \langle 0, 1, 1/3 \rangle\}.$$

Remark 4.4.2. *Note that we can get bases for the row, column, and null space of a matrix by way of its reduced echelon form. But we use the echelon form in different ways for each of these tasks.*

- *To obtain a basis for $\mathcal{N}(A)$, we use the entries in* $\text{rref}(A)$ *to deduce the relationship between basic and free variables. We use these to characterize solutions to $A\vec{x} = \vec{0}_m$.*

- *To obtain a basis for $\mathcal{CS}(A)$, we use use* $\text{rref}(A)$ *to identify the pivot columns, and we take the pivot columns from A to form our basis elements.*

- *To obtain a basis for $\mathcal{RS}(A)$, we take the nonzero rows of* $\text{rref}(A)$ *as our basis elements.*

Caveat: *If the first k rows of* $\mathrm{rref}(A)$ *are nonzero, it is not necessarily true that the first k rows of A are linearly independent! This makes sense given that we're allowed to move the rows around. Hence when obtaining a basis for $\mathcal{RS}(A)$, we take the row vectors from the echelon form, not the original matrix.*

We can summarize the relationships between row operations and the row and column spaces of a matrix.

- Elementary row operations preserve the row space of a matrix but may change the linear dependence relations between the rows.
- Elementary row operations preserve the linear dependence relation between the columns but may change the column space.

The punch line of these observations is that we use columns of the original matrix for a column space basis and rows of the echelon form for a row space basis.

Example 4.4.4. *Find bases for $\mathcal{RS}(A)$, $\mathcal{CS}(A)$, and $\mathcal{N}(A)$, where*

$$A = \begin{bmatrix} 1 & -2 & 0 & 1 & 8 \\ 2 & -4 & 3 & 0 & -6 \\ -1 & 2 & 1 & 2 & 3 \\ 3 & -6 & 4 & 0 & -7 \end{bmatrix}.$$

we can get all three from $\mathrm{rref}(A)$. *Note that*

$$\mathrm{rref}(A) = \begin{bmatrix} 1 & -2 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & -4 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

From the rref, we see that for the equation $A\vec{x} = \vec{0}_4$, the solution $\vec{x} \in R^5$ will have three basic and two free variables. The solution to this homogeneous equation can be expressed in vector parametric form

$$\vec{x} = s\langle 2, 1, 0, 0, 0\rangle + t\langle -3, 0, 4, -5, 1\rangle.$$

We also see that the pivot columns are the first, third and fourth, so we can take these columns of A as the basis elements of the columns space. The basis *elements for the row space will be the three nonzero rows of* rref(A)*. We have the following solution where each spanning set shown is a basis:*

$$\begin{aligned} \mathcal{RS}(A) &= \text{Span}\left\{\langle 1,-2,0,0,3\rangle, \langle 0,0,1,0,-4\rangle, \langle 0,0,0,1,5\rangle\right\}, \\ \mathcal{CS}(A) &= \text{Span}\left\{\langle 1,2,-1,3\rangle, \langle 0,3,1,4\rangle, \langle 1,0,2,0\rangle\right\}, \\ \mathcal{N}(A) &= \text{Span}\left\{\langle 2,1,0,0,0\rangle, \langle -3,0,4,-5,1\rangle\right\}. \end{aligned}$$

Exercise 4.4.2. *Find bases for the row space, column space, and null space of each matrix.*

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

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

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

4.5 The Fundamental Theorem of Linear Algebra

Now that we have a method for finding bases of the fundamental subspaces of a matrix, we are able to deduce their dimensions. In particular, given a matrix A we have a connection between the number of its pivot columns and the dimensions of each of $\mathcal{CS}(A)$, $\mathcal{RS}(A)$ and $\mathcal{N}(A)$. Theorem 4.4.1 tells us that the pivot columns form a basis for $\mathcal{CS}(A)$, hence

$$\dim\big(\mathcal{CS}(A)\big) = \text{the number of pivot columns of } A. \tag{4.22}$$

For the null space, if the homogeneous equation $A\vec{x} = \vec{0}_m$ has only the trivial solution, then $\mathcal{N}(A) = \left\{\vec{0}_m\right\}$. In this case, $\dim(\mathcal{N}(A)) = 0$ by definition. If $A\vec{x} = \vec{0}_m$ permits nontrivial solutions, then our method of decomposing a solution of $A\vec{x} = \vec{0}_m$ according to the free variables produces a basis for

$\mathcal{N}(A)$. The number of vectors in that basis is the number of free variables. Recalling that free variable correspond to non-pivot columns, we have

$$\dim\big(\mathcal{N}(A)\big) = \text{the number of non-pivot columns of } A. \tag{4.23}$$

The set of nonzero row vectors of $\text{rref}(A)$ form a basis for $\mathcal{RS}(A)$. Let's recall that every pivot position in a matrix A occupies a row and a column. So the number of rows with a pivot position (i.e., nonzero rows) necessarily is the same as the number of pivot columns. After all, the leftmost nonzero entry in each nonzero row of $\text{rref}(A)$ is one of the leading ones. It follows that

$$\dim\big(\mathcal{RS}(A)\big) = \text{the number of pivot columns of } A. \tag{4.24}$$

If A is an $m\times n$ matrix, then $\mathcal{CS}(A)$ is a subspace of R^m whereas $\mathcal{RS}(A)$ is a subspace of R^n. So, except in the case of a square matrix, these two spaces contain different sorts of vectors. In Example 4.4.4, for example, the basis elements we found for $\mathcal{CS}(A)$ have four entries each while the basis elements for $\mathcal{RS}(A)$ have five entries each. This makes sense given that A was 4×5. But even though the types of vectors are not comparable between these two spaces, the number of elements in their bases turned out to be the same. It turns out that this is not a coincidence or novel feature of this example, but rather is related to a property of matrices in general—for any matrix A, $\dim\big(\mathcal{CS}(A)\big) = \dim\big(\mathcal{RS}(A)\big)$. This number is given a special name; it's called a *rank*.

Definition 4.5.1. *The* **rank** *of a matrix A, denoted* $\text{rank}(A)$*, is the dimension of the column space of A.*

We also have a special name for the dimension of the null space of a matrix. We call this the nullity.

Definition 4.5.2. *The* **nullity** *of a matrix A, denoted* $\text{nullity}(A)$*, is the dimension of the null space of A.*

We are ready to state the Fundamental Theorem of Linear Algebra, the proof of which has been argued (in pieces) in sections 4.2.1, 4.4.1, 4.4.2, and the current section.

Theorem 4.5.1. The Fundamental Theorem of Linear Algebra
Let A be an $m \times n$ matrix. Then

1. $\text{rank}(A) = \dim\big(\mathcal{CS}(A)\big) = \dim\big(\mathcal{RS}(A)\big)$.

2. $\text{rank}(A) + \text{nullity}(A) = n$.

3. *Every vector $\vec{x}$ in $\mathcal{RS}(A)$ is orthogonal to every vector $\vec{y}$ in $\mathcal{N}(A)$, and similarly, every vector $\vec{u}$ in $\mathcal{CS}(A)$ is orthogonal to every vector $\vec{v}$ in $\mathcal{N}(A^T)$.*

Part 2. of Theorem 4.5.1 is sometimes referred to as the "**rank-nullity theorem**." It seems rather surprising at first glance—the rank of a matrix plus the dimension of its null space must equal its number of columns. But equations (4.22) and (4.23) eliminate any sense of mystery about this result. If the matrix A has n columns, given that every column is either a pivot column or a non-pivot column, part 2 can be restated as

$$\begin{array}{rl} & \text{the number of pivot columns of } A \\ + & \text{the number of non-pivot columns of } A \\ \hline = & \text{the total number of columns of } A. \end{array}$$

This is rather obvious!

Example 4.5.1. *Suppose A is a 6×14 matrix.*

1. *If* $\text{rank}(A) = 5$*, determine* $\dim(\mathcal{RS}(A))$.

2. *If* $\text{rank}(A) = 4$*, determine* $\text{nullity}(A)$.

3. *What is the maximum possible rank of A?*

4. *If* $\text{rank}(A) = 3$*, what is* $\text{nullity}(A^T)$*?*

5. *If* $\text{nullity}(A) = 12$*, what is the dimension of the row space of A?*

We can answer each question by applying and appropriate part of Theorem 4.5.1. For question 1., the first part of the theorem says that the rank and dimension of the row space are the same. So

$$\dim(\mathcal{RS}(A)) = 5.$$

For question 2., we can apply the rank-nullity theorem (part 2 of Theorem 4.5.1), $\text{rank}(A) + \text{nullity}(A) = n$*, so that*

$$4 + \text{nullity}(A) = 14 \quad \textit{making} \quad \text{nullity}(A) = 10.$$

For question 3., we note that there are six rows so that the maximum number of pivot positions is six. So the maximum value of the rank is 6. *That is,*

$$\mathrm{rank}(A) \leq 6.$$

For question 3., we note that A^T *is a* 14×6 *matrix. So the new* n *value for the rank-nullity theorem would be* 6. *Since*

$$\mathrm{rank}(A^T) = \dim(\mathcal{CS}(A^T)) = \dim(\mathcal{RS}(A)) = \mathrm{rank}(A),$$

we have

$$3 + \mathrm{nullity}(A^T) = 6 \quad \textit{making} \quad \mathrm{nullity}(A^T) = 3.$$

Finally, to answer part 5., we can use that $\dim(\mathcal{RS}(A)) = \mathrm{rank}(A)$. *Applying the rank-nullity theorem, we have*

$$\dim(\mathcal{RS}(A)) + 12 = 14 \quad \textit{giving} \quad \dim(\mathcal{RS}(A)) = 2.$$

Exercise 4.5.1. *Suppose* A *is a* 10×20 *matrix.*

1. *If* $\mathrm{rank}(A) = 7$, *what is* $\mathrm{nullity}(A)$*?*
2. *If* $\mathrm{rank}(A) = 7$, *what is* $\mathrm{nullity}(A^T)$*?*
3. *If* A^T *has* 8 *pivot columns, what is* $\mathrm{rank}(A)$*?*
4. *If* $\dim\big(\mathcal{CS}(A^T)\big) = 9$, *how many free variables are there in any solution to* $A\vec{x} = \vec{0}_{10}$*?*
5. *What is the maximum possible rank of* A*?*

Exercise 4.5.2. *Explain why the maximum rank of an* $m \times n$ *matrix is the smaller of the two numbers,* m *and* n.

Remark 4.5.1. *A matrix that has the maximum rank that it can have for its size is often called* **full rank**. *For example, an* $m \times n$ *matrix* A *is considered full rank if* $\mathrm{rank}(A) = \min\{m, n\}$.

Exercise 4.5.3. *Suppose* A *is an* $n \times n$ *matrix (so* A *is square). Explain why if* A *is full rank, then* $\mathrm{nullity}(A) = 0$.

As a historical note, the name "Fundamental Theorem of Linear Algebra" was introduced by Gilbert Strang, a prominent mathematician at M.I.T., in his 1988 textbook *Linear Algebra and Its Applications* [1] and in a 1993 article in *The American Mathematical Monthly* [2]. Strang, who devoted much of his career to studying and teaching linear algebra at M.I.T. viewed the compilation of facts in Theorem 4.5.1 as being appropriate to be given the "fundamental theorem" designation for the subject of linear algebra.[6] Strang observed that other major fields of mathematics (such as calculus, number theory, and abstract algebra) already had theorems that are designated as being fundamental, and believed that linear algebra should also have such a theorem. A theorem that is designated as "fundamental" in any mathematical subject should be a theorem that encapsulates and synthesizes some major (and non-trivial) facts from the subject in a nutshell. The Fundamental Theorem of Calculus is the perfect example. The main topics of calculus are limits, derivatives and integrals, and the Fundamental Theorem of Calculus relates and synthesizes these major concepts in the form of a single theorem. The Fundamental Theorem of Calculus can't be introduced until the end of a Calculus I course because there is much preliminary work (the study of limits, derivatives, and integrals) that needs to take place before the student can understand the fundamental theorem that relates these concepts. When the theorem is finally introduced, it is seen to be an elegant synthesis of all that was studied in the Calculus I course. Likewise, there is much preliminary work that we need to do in a linear algebra course before Theorem 4.5.1 can be understood. We need to understand the basics of vectors, the concept of orthogonality, the concept of subspace, and the concept of dimension before we can understand Theorem 4.5.1. Once we do understand these concepts, then Theorem 4.5.1 is seen to tie the concepts together into one cohesive package.

At the present time, the name "Fundamental Theorem of Linear Algebra" is not used in most linear algebra textbooks in reference to the facts in Theorem 4.5.1, but the name seems to be catching on (as you will see if you search the internet). We have chosen to use the name in this text, in agreement with Strang's thinking that if linear algebra is to have a "fundamental theorem" then Theorem 4.5.1 aptly fits that role. It should be noted that Theorem 4.5.1 is the most basic statement of the fundamental theorem. There is a more comprehensive version of the theorem that includes informa-

[6] Gilbert Strang retired from M.I.T in 2023.

tion how to find the most useful bases for each of the fundamental subspaces. We will introduce this more complete version of the theorem after we have introduced the needed relevant concepts in Chapter 7.

4.6 General Vector Spaces

The phrase *vector space* has been used throughout this text without an attempt to carefully define it. That lack of a formal definition has not interfered with our ability to explore R^n, to use matrices and vectors in R^n to solve and express solutions of systems of linear equations, and to perform algebra on matrices. In this section, we will step back and generalize the sort of structure that we see with R^n. Let's recall the basic framework of R^n and consider the algebraic properties associated with the two key operations, vector addition and scalar multiplication.

We defined vectors in the vector space R^n as ordered n-tuples of real numbers, and together with scalars, we defined the operations of vector addition and scalar multiplication. These operations performed on vectors in R^n produce vectors in R^n and satisfy some algebraic properties. In particular, if $\vec{x}$, $\vec{y}$, and $\vec{z}$ are any vectors in R^n and c and d are any scalars, then

- The vector $\vec{x} + \vec{y}$ is in R^n,
- the vector $c\vec{x}$ is in R^n,
- $\vec{x} + \vec{y} = \vec{y} + \vec{x}$,
- $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$,
- there is a vector $\vec{0}_n$ such that $\vec{x} + \vec{0}_n = \vec{x}$,
- there is a vector $-\vec{x}$ such that $-\vec{x} + \vec{x} = \vec{0}_n$,
- $c(\vec{x} + \vec{y}) = c\vec{x} + c\vec{y}$,
- $(c + d)\vec{x} = c\vec{x} + d\vec{x}$,
- $c(d\vec{x}) = (cd)\vec{x} = d(c\vec{x})$, and
- $1\vec{x} = \vec{x}$.

Some of these properties were stated explicitly, while others may have been taken for granted as following from our knowledge of addition and multiplication of real numbers. A **real vector space** is an abstraction of what we see with R^n. As with R^n, vectors are accompanied by scalars. The word "real" in the phrase "real vector space" tells us that the set of scalars[7] will be the set of real numbers, R. Since we won't consider other types of scalars, we will drop the word *real* following the definition and just use the phrase vector space.

Definition 4.6.1. *A* ***real vector space*** *is a set, V, of objects called vectors together with two operations called* ***vector addition*** *and* ***scalar multiplication*** *that satisfy the following axioms:*

For each vector $\vec{x}$, $\vec{y}$, and $\vec{z}$ in V and for any scalars, c and d

1. *the sum $\vec{x}+\vec{y}$ is in V, and*
2. *the scalar multiple $c\vec{x}$ is in V.*
3. $\vec{x}+\vec{y}=\vec{y}+\vec{x}$,
4. $(\vec{x}+\vec{y})+\vec{z}=\vec{x}+(\vec{y}+\vec{z})$,
5. *There is an additive identity vector in V called the zero vector denoted $\vec{0}$, such that $\vec{x}+\vec{0}=\vec{x}$ for every $\vec{x}$ in V,*
6. *For each vector $\vec{x}$ in V, there is an additive inverse vector denoted $-\vec{x}$ such that $-\vec{x}+\vec{x}=\vec{0}$.*
7. $c(\vec{x}+\vec{y})=c\vec{x}+c\vec{y}$,
8. $(c+d)\vec{x}=c\vec{x}+d\vec{x}$,
9. $c(d\vec{x})=(cd)\vec{x}=d(c\vec{x})$, *and*
10. $1\vec{x}=\vec{x}$.

Remark 4.6.1. *Until now, the term* ***vector*** *has exclusively been used to refer to an element of R^n, an ordered n-tuple characterized by a magnitude and direction. Now, the term* ***vector*** *can be used to refer to an element of any vector space.*

Remark 4.6.2. *An* ***axiom*** *is a statement that is specified as being true. So an axiom does not require proof. If it has been established that some set is a vector space, then each of the properties in Definition 4.6.1 are automatically known to hold. If we have a set of objects with operations that might be a vector space, then the properties in Definition 4.6.1 can be used to test whether the set really is a vector space.*

[7] While we won't use alternatives here, there are sets of scalars other than R.

We recognize the properties in axioms 1 and 2 from Definition 4.2.1 of a subspace of R^n. We called this property *being closed*. So a vector space is closed with respect to the two operations of vector addition and scalar multiplication. Axioms 5 and 6 specify that there is an additive identity vector and each vector has an additive inverse, while axiom 10 specifies that the scalar 1 is the scalar multiplicative identity. In light of Definition 4.6.1, R^n is a real vector space. In general, the objects in a vector space could be anything—provided we can define the operations so that the necessary properties hold. Note that it is not sufficient to simply specify what the objects are when defining a vector space. The definition of a specific vector space must include a description of the two operations.

A very simple example of a vector space is the *trivial vector space* which consists of just one vector. It doesn't matter what we call the vector but it makes sense to call the vector $\vec{0}$. This is because we know that any vector space must have a zero vector, and if there is only one vector in the vector space, then it must be a zero vector. So this vector space is $V = \left\{\vec{0}\right\}$ and we define the operations of vector addition and scalar multiplication in this vector space by

$$\vec{0} + \vec{0} = \vec{0}$$

and for any scalar c,

$$c\vec{0} = \vec{0}.$$

$V = \left\{\vec{0}\right\}$ with operations as defined above satisfies all of the axioms given in Definition 4.6.1. Thus V is a vector space that consists of only one vector.

Exercise 4.6.1. *Verify that $V = \left\{\vec{0}\right\}$ with operations as defined above satisfies all of the axioms given in Definition 4.6.1 (and is thus a vector space).*

Exercise 4.6.2. *Are there any vector spaces that have exactly two elements? Explain why or why not.*

Hint to get started thinking about this problem: Suppose that we have a vector space V that contains exactly two elements. One of these elements must be a zero vector and one of them must be something else, since we are assuming there are two vectors in the vector space. We can call the zero vector $\vec{0}$ and call the other vector $\vec{v}$. So we have a vector space $V = \left\{\vec{0}, \vec{v}\right\}$ where $\vec{v} \neq \vec{0}$. Now see what you can deduce using the axioms of Definition 4.6.1.

There are some properties of the vector spaces R^n that we have perhaps taken for granted:

1. There is only one additive identity vector in R^n. (i.e., the zero vector in R^n is unique.)

2. Every vector in R^n has exactly one additive inverse. (i.e., the additive inverse of any vector in R^n is unique.)

3. If $\vec{x}$ is any vector in R^n, then $0\vec{x} = \vec{0}_n$.

4. If c is any scalar, then $c\vec{0}_n = \vec{0}_n$.

Although the above four properties seem to be obvious, let us write proofs of these facts.

Proof of 1: A vector $\vec{y}$ is said to be an additive identity for R^n if it is true that $\vec{x} + \vec{y} = \vec{x}$ for all vectors $\vec{x}$ in R^n. We know that the vector $\vec{0}_n = \langle 0, 0, \ldots, 0 \rangle$ serves as an additive identity for R^n because it is easy to see that $\vec{x} + \vec{0}_n = \vec{x}$ for all vectors $\vec{x}$ in R^n. But might there be some other vector $\vec{y} = \langle y_1, y_2, \ldots, y_n \rangle$ in R^n that serves as an additive identity for R^n? If so then $\vec{x} + \vec{y} = \vec{x}$ for all vectors $\vec{x}$ in R^n and in particular $\vec{y} + \vec{y} = \vec{y}$. Written out in long form

$$\langle y_1, y_2, \ldots, y_n \rangle + \langle y_1, y_2, \ldots, y_n \rangle = \langle y_1, y_2, \ldots, y_n \rangle ,$$

and by our definition of vector addition in R^n,

$$\langle 2y_1, 2y_2, \ldots, 2y_n \rangle = \langle y_1, y_2, \ldots, y_n \rangle .$$

By equating the first components on each side of the above equation, we obtain $2y_1 = y_1$, but this can be true if and only if $y_1 = 0$. By equating the other components, we see that $y_i = 0$ for all $i = 1, 2, \ldots, n$. Thus it must be the case that $\vec{y} = \vec{0}_n$. We have proved that there is only one additive identity element in R^n. It is the zero vector, $\vec{0}_n = \langle 0, 0, \ldots, 0 \rangle$.

Proof of 2: Next we will prove that every vector in R^n has exactly one additive inverse. An additive inverse of a vector $\vec{x}$ in R^n is a vector $\vec{y}$ in R^n such that $\vec{x} + \vec{y} = \vec{0}_n$. Suppose that $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle$ is a given vector in R^n. We know that the vector $-\vec{x} = \langle -x_1, -x_2, \ldots, -x_n \rangle$ serves as an additive inverse for $\vec{x}$ because $\vec{x} + (-\vec{x}) = \vec{0}_n$. But might there be some other vector $\vec{y} = \langle y_1, y_2, \ldots, y_n \rangle$ such that $\vec{x} + \vec{y} = \vec{0}_n$? If so, then

$$\langle x_1, x_2, \ldots, x_n \rangle + \langle y_1, y_2, \ldots, y_n \rangle = \langle 0, 0, \ldots, 0 \rangle .$$

By using our definition of vector addition in R^n, we obtain

$$\langle x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n \rangle = \langle 0, 0, \ldots, 0 \rangle .$$

Equating the first components in the above equation, we obtain $x_1 + y_1 = 0$ which implies that $y_1 = -x_1$. Equating all of the other components, we obtain $y_i = -x_i$ for all $i = 1, 2, \ldots, n$ and thus

$$\vec{y} = \langle y_1, y_2, \ldots, y_n \rangle = \langle -x_1, -x_2, \ldots, -x_n \rangle = -\vec{x}.$$

We have proved that each vector $\vec{x}$ in R^n has exactly one additive inverse. It is $-\vec{x} = \langle -x_1, -x_2, \ldots, -x_n \rangle$.

Proof of 3: If $\vec{x} = \langle x_1, x_2, \ldots x_n \rangle$ is any vector in R^n, then

$$0\vec{x} = 0 \langle x_1, x_2, \ldots x_n \rangle = \langle 0x_1, 0x_2, \ldots 0x_n \rangle = \langle 0, 0, \ldots, 0 \rangle = \vec{0}.$$

Proof of 4: If c is any scalar, then

$$c\vec{0} = c \langle 0, 0, \ldots, 0 \rangle = \langle 0, 0, \ldots, 0 \rangle = \vec{0}.$$

The analogues of all four of the above facts (1, 2, 3, and 4) are true in any vector space. However, to prove them in an arbitrary vector space, we need to do it by only using the vector space axioms given in Definition 4.6.1. When thinking of a general vector space, we are not allowed to assume that the vectors in the vector space have any specific form. In particular, we can't assume that the vectors are ordered n–tuples of real numbers, as they are in R^n. The theorem below lists the four basic properties of vector spaces that are the analogues of facts 1, 2, 3, and 4 given above.

Theorem 4.6.1. *Suppose that V is a vector space. Then*

1. *There is only one additive identity vector in V (i.e., the zero vector of V is unique).*

2. *Each vector in V has only one additive inverse (i.e., the additive inverse of any vector in V is unique).*

3. *If $\vec{x}$ is any vector in V, then $0\vec{x} = \vec{0}$.*

4. *If c is any scalar, then $c\vec{0} = \vec{0}$.*

We will prove statements 1 and 3 above and leave the proofs of statements 2 and 4 as exercises. The thing that you should pay attention to when reading these proofs is that we do not assign any particular form to vectors. In particular, we do not assume that vectors are ordered n–tuples of numbers. We only use the vector space axioms that are given in Definition 4.6.1.

Proof. First we will prove statement 1, which says that the zero vector of V is unique. By vector space axiom 5 of Definition 4.6.1, we know that an additive identity vector exists in V (by assumption). Suppose that there are two additive identity vectors in V, which we can call $\vec{y}$ and $\vec{z}$. Since these are additive identity vectors, we have

$$\vec{x}+\vec{y}=\vec{x} \quad \text{and} \quad \vec{x}+\vec{z}=\vec{x}$$

for all vectors $\vec{x}$ in V. In particular, since $\vec{y}$ is an additive identity and $\vec{z}$ is in V,

$$\vec{z}+\vec{y}=\vec{z}.$$

On the other hand, since $\vec{z}$ is an additive identity and $\vec{y}$ is in V,

$$\vec{y}+\vec{z}=\vec{y}.$$

Vector space axiom 3 says that addition is commutative, that is $\vec{z}+\vec{y}=\vec{y}+\vec{z}$. So these two equations imply that $\vec{z}=\vec{y}$. Therefore there is only one additive identity vector in V, and this completes the proof of statement 1. □

In stating vector space axiom 5, we gave the name $\vec{0}$ to this additive identity vector. Now that we have proved that $\vec{0}$ is unique, we can refer to it as **the** additive identity element (or **the** zero vector) of V, rather than saying **an** additive identity element, as we did when stating the axiom.

Now we will prove statment 3. Again, if we knew that we were working in R^n, then proving statement 3 would be easy, but we are not assuming that we are working in R^n, so we can only use the vector space axioms in our proof.

Proof. We want to prove that if $\vec{x}$ is any vector in V, then $0\vec{x}=\vec{0}$. To do this, we begin by letting $\vec{x}$ be any vector in V. By vector space axiom 2 (closure under scalar multiplication), we know that $0\vec{x}$ is also in V. Next we note that $0+0=0$ (a well known fact about scalars), and thus

$$(0+0)\,\vec{x}=0\vec{x}.$$

By vector space axiom 8, we have

$$0\vec{x} + 0\vec{x} = (0 + 0)\,\vec{x},$$

and thus

$$0\vec{x} + 0\vec{x} = 0\vec{x}.$$

By vector space axiom 6, we know that the vector $0\vec{x}$ has an additive inverse in V. We call it $-(0\vec{x})$. Adding this additive inverse to both sides of the above equation gives

$$(0\vec{x} + 0\vec{x}) + (-(0\vec{x})) = 0\vec{x} + (-(0\vec{x}))\,.$$

Using vector space axiom 4 (associativity), we can rewrite the above equation as

$$0\vec{x} + (0\vec{x} + (-(0\vec{x}))) = 0\vec{x} + (-(0\vec{x}))\,.$$

However, $0\vec{x} + (-(0\vec{x})) = \vec{0}$ and thus we have

$$0\vec{x} + \vec{0} = \vec{0}.$$

Finally, note that $0\vec{x} + \vec{0} = 0\vec{x}$ by axiom 5, and thus we have

$$0\vec{x} = \vec{0},$$

which is what we wanted to prove. □

Our experience working in R and in R^n may lead us to think of a property like $0\vec{x} = \vec{0}$ as somehow *obvious* or self-evident. But without that specific context, it is very interesting to see how many of the vector space axioms are needed to prove this seemingly simple property.

Exercise 4.6.3. *Prove statements 2 and 4 of Theorem 4.6.1.*

Exercise 4.6.4. *We know by statement 2 of Theorem 4.6.1 that if V is a vector space and $\vec{x}$ is any vector in V, then the additive inverse of $\vec{x}$ is unique. We give this vector the name $-\vec{x}$. Prove that if $\vec{x}$ is any vector in V, then $-\vec{x} = (-1)\,\vec{x}$.*

Note that this is very easy to prove in R^n because in R^n, if we have $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle$, then

$$-\vec{x} = \langle -x_1, -x_2, \ldots, -x_n \rangle$$

and

$$(-1)\,\vec{x} = -1\,\langle x_1, x_2, \ldots, x_n\rangle = \langle -x_1, -x_2, \ldots, -x_n\rangle\,.$$

However, in doing this exercise you may not assume that $\vec{x}$ is in R^n! You can only use the vector space axioms (and any of the facts from Theorem 4.6.1, now that the theorem has been proven).

Some of the most important ideas we have studied regarding the vector spaces R^n are the ideas of linear combinations, spans, linear independence, bases, dimension, and coordinate vectors. Now that we have defined what a vector space is in a broader sense, we can immediately generalize all of these ideas in regard to general vector spaces. The definitions given below should look familiar to you because you have already seen them stated in the setting of R^n earlier in this chapter. The only difference is that we are stating them with respect to any vector space (not just R^n).

Definition 4.6.2. *Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ be a set of one or more ($k \geq 1$) vectors in a vector space V. A* ***linear combination*** *of these vectors is any vector of the form*

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k,$$

where $x_1, , x_2, \ldots, x_k$ are scalars. The coefficients, $x_1, x_2, \ldots, x_k$, are often called the ***weights****.*

Definition 4.6.3. *Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ be a set of one or more ($k \geq 1$) vectors in a vector space V. The set of all possible linear combinations of the vectors in S is called the* ***span*** *of S. It is denoted by* $\text{Span}(S)$ *or by* $\text{Span}\,\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$.

Definition 4.6.4. *Let V be a vector space. The collection of vectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ in V is said to be* ***linearly independent*** *if the homogeneous equation*

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k = \vec{0} \tag{4.25}$$

has only the trivial solution, $x_1 = x_2 = \cdots = x_k = 0$. A set of vectors that is not linearly independent is called ***linearly dependent****.*

Definition 4.6.5. *Let V be a real vector space. A* ***subspace*** *of V is a nonempty set, S, of vectors in V such that*

- *for every $\vec{x}$ and $\vec{y}$ in S, $\vec{x} + \vec{y}$ is in S, and*

- *for each $\vec{x}$ in S and scalar c, $c\vec{x}$ is in S.*

Definition 4.6.6. *Let S be a subspace of a vector space V, and let $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_k\}$ be a subset of vectors in S. $\mathcal{B}$ is a **basis** of S provided*

- $\text{Span}(\mathcal{B}) = S$
- *$\mathcal{B}$ is linearly independent.*

We have learned that if S is any subspace of R^n that has a basis containing exactly k vectors, then any basis of S must contain exactly k vectors. We used this fact to define the dimension of S to be $\dim(S) = k$. The analogue of this fact is true (and its proof is similar) in any vector space V. Specifically, if S is any subspace of V that has a basis containing exactly k vectors, then any basis of S must contain exactly k vectors. It thus makes sense to define dimension of S to be $\dim(S) = k$.

In R^n, the trivial subspace $S = \left\{\vec{0}_n\right\}$ does not have a basis and we define its dimension to be 0. If V is any vector space, then V contains the trivial subspace $S = \left\{\vec{0}_n\right\}$, which does not have a basis, and it clearly makes sense to define the dimension of this subspace to be 0.

Any vector space (or subspace of a vector space) that has a basis containing a finite number of vectors is called a *finite dimensional* vector space and its dimension is defined to be the number of vectors in any of its bases. (A trivial vector space is also said to be finite dimensional and its dimension is defined to be 0.) All of the vector spaces R^n and all subspaces of R^n are finite dimensional. A situation that has not arisen in our study of R^n is the fact that a vector space might not have a basis that consists of a finite number of vectors. Such vector spaces are said to be *infinite dimensional.* These are, in fact, the most interesting vector spaces from the point of view of mathematical analysis and they are the setting in which the most powerful applications of linear algebra are seen. We will soon see some examples of infinite dimensional vector spaces, but an in–depth study of them will not be undertaken in this course. We will however study some finite dimensional subspaces of these infinite dimensional vector spaces.

Having discussed these issues regarding dimensions, let us summarize with the following definition.

Definition 4.6.7. *Let S be a subspace of a vector space V. We define the **dimension** of S as follows:*

- *If* $S = \{\vec{0}\}$, *then we define* $\dim(S) = 0$.
- *If* S *has a basis consisting of* k *vectors, where* $k < \infty$, *then we define* $\dim(S) = k$.
- *If* S *is not spanned by any finite set of vectors, then we say that* S *is infinite dimensional.*

Before proceeding to look at some examples of vector spaces other than R^n, let us recall one more important idea concerning R^n – the idea of coordinate vectors. If S is a finite dimensional subspace of R^n and $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_k\}$ is an ordered basis for S, then for any vector $\vec{x}$ in S, there is a *unique* set of scalars $c_1, c_2, \ldots, c_k$ such that $\vec{x} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k$. The vector $[\vec{x}]_{\mathcal{B}} = \langle c_1, c_2, \ldots, c_k \rangle$, which is a vector in R^k, is called the coordinate vector $\vec{x}$ with respect to the basis $\mathcal{B}$. It is likewise true for any finite dimensional subspace, S, of any vector space V, that if $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_k\}$ is an ordered basis for S and $\vec{x}$ in S, then there is a unique set of scalars $c_1, c_2, \ldots, c_k$ such that $\vec{x} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k$. It thus makes sense to make the following definition.

Definition 4.6.8. *Suppose that* V *is a vector space and suppose that* S *is a finite dimensional subspace of* V. *Suppose that* $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_k\}$ *is an ordered basis for* S. *Then the (unique) vector*

$$[\vec{x}]_{\mathcal{B}} = \langle c_1, c_2, \ldots, c_k \rangle \in R^k$$

such that

$$\vec{x} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k$$

is called the ***coordinate vector*** *of* $\vec{x}$ *with respect to the ordered basis* $\mathcal{B}$.

In Section 4.8, we provide further discussion on how coordinate vectors can be employed in working with finite dimensional subspaces of any vector space.

4.7 Examples of Vector Spaces

We will now look at some examples of vector spaces other than R^n. In looking at these examples, we will highlight the big ideas that have been defined above – subspace, span, linear independence, basis, dimension, and coordinate vectors. The objects that make up the vector spaces in the examples that we will give are not vectors in R^n, but they are objects that we are familiar with. Our first example (Section 4.7.2) will be of vector spaces whose vectors are *matrices.* This example serves as a gentle introduction to general vector spaces, since operating with matrices is very similar to operating with vectors in R^n. Our second example (Section 4.7.3) is of the vector space R^∞, which is a natural generalization of R^n, but with the big difference that it is an infinite dimensional vector space. Its elements are *infinite sequences* of real numbers. Our third collection of examples (Section 4.7.4) will be of vector spaces whose vectors are *functions.* These vector spaces are often referred to as *function spaces* and they provide the setting in which the tools and theory of linear algebra can be employed in other areas of mathematics, such as calculus, differential equations, and functional analysis.

4.7.1 A Note on Notation

We have been using the arrow notation ($\vec{x}$, $\vec{y}$, $\vec{v}$, etc.) throughout our discussion of vector spaces, their basic properties, and basic definitions that apply to their study. This is good practice when we are having general discussions about vector spaces (with no particular vector space in mind) because it helps us to keep our thinking straight about whether we are dealing with a vector or a scalar. When we write $\vec{x}$ (with an arrow over it), we know we are referring to a vector. When we write c (with no arrow over it), we know we are referring to a scalar. That having been said, when we are in a specific setting in which the vector space we are studying consists of some objects that we are familiar to us, and which do not normally have any "arrow" notation associated with them, we do not use an arrow notation, even though the objects are the vectors of the vector space under consideration. For example, we are going to look at examples of vector spaces whose vectors are matrices (in Section 4.7.2). We are accustomed to denoting a matrix using a capital letter such as A. We are not accustomed to writing $\vec{A}$, and hence we will not do that, even though we want to consider A to be a vector. Likewise, we will be considering vector spaces whose vectors are functions (in Section 4.7.4) . In calculus, it is customary to name functions using lower case letters such as f and g. We will continue to adopt that convention, rather than writing $\vec{f}$ and $\vec{g}$, even though these functions will be the vectors of our vector space.

4.7.2 Vector Spaces of Matrices

In Chapter 3, we saw that matrices could be thought of as objects that we can manipulate with algebraic operations. In fact, the first two operations defined in Section 3.2 were addition and scalar multiplication of matrices. It thus probably comes as no surprise that there are vector spaces whose vectors are matrices. Let $M_{m\times n}$ denote the set of all $m \times n$ matrices with real entries. We define vector addition and scalar multiplication as regular addition of matrices and scalar multiplication of matrices (as defined in Section 3.2). $M_{m\times n}$ is a real vector space.

Example 4.7.1. *What is the zero vector in* $M_{m\times n}$*?*

As you probably guessed, the zero vector in $M_{m\times n}$ *is the* $m\times n$ *zero matrix,* $O_{m\times n}$*. This is the* $m \times n$ *matrix all of whose entries are* 0*. If we take any* $m \times n$ *matrix, then* $A + O_{m\times n} = A$*. Also, recall that statement 1 of Theorem 4.6.1 says that the zero vector in any vector space is unique. So there is no matrix other than* $O_{m\times n}$ *that serves as an additive identity in the vector space* $M_{m\times n}$*.*

Exercise 4.7.1. *Let*

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

be any element of $M_{2\times 2}$*. Show that* $-1A$ *is the additive inverse of* A*.*

Exercise 4.7.2. *Show that the matrix*

$$A = \begin{bmatrix} 5 & -5 \\ -1 & 4 \end{bmatrix}$$

in $M_{2\times 2}$ *is a linear combination of the matrices in the set of matrices* $S = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ *where*

$$E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},\ E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},\ E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Exercise 4.7.3. *Show that any matrix*

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in M_{2\times 2}$$

is a linear combination of the matrices $S = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ *given in Exercise 4.7.2.*

In other words, show that $\mathrm{Span}\,(S) = M_{2\times 2}$*.*

Exercise 4.7.4. *Show that the set of matrices* $S = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ *given in Exercise 4.7.2 is linearly independent. To do this, you will need to show that the equation*

$$x_{11}E_{11} + x_{12}E_{12} + x_{21}E_{21} + x_{22}E_{22} = O_{2\times 2}$$

has only the trivial solution $(x_{11} = x_{12} = x_{21} = x_{22} = 0)$.

Fill in the blank: Since the set $S = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ *is linearly independent and* $\operatorname{Span}(S) = M_{2\times 2}$*, then* S *is a* ______ *for* $M_{2\times 2}$.

What is the dimension of $M_{2\times 2}$*?*

Exercise 4.7.5. *Consider the ordered basis* $\mathcal{B} = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ *given in Exercise 4.7.2.*

1. *If* A *is any matrix in* $M_{2\times 2}$*, then the coordinate vector* $[A]_{\mathcal{B}}$ *is a vector in* R^k*. Determine the value of* k.

2. *Find the coordinate vectors* $[A]_{\mathcal{B}}$ *and* $[B]_{\mathcal{B}}$ *for the matrices*

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

3. *Evaluate* $A + B$ *and confirm that* $[A]_{\mathcal{B}} + [B]_{\mathcal{B}} = [A + B]_{\mathcal{B}}$

4. *Evaluate* $5A$ *and confirm that* $5[A]_{\mathcal{B}} = [5A]_{\mathcal{B}}$.

5. *Find the coordinate vectors for the elements of* $\mathcal{B}$*. That is, find each of* $[E_{11}]_{\mathcal{B}}$, $[E_{12}]_{\mathcal{B}}$, $[E_{21}]_{\mathcal{B}}$*, and* $[E_{22}]_{\mathcal{B}}$.

6. *Can you make a conjecture about what the coordinate vectors should be for the basis elements of a basis in general?*

Exercise 4.7.6. *Find a set of matrices that is a basis for* $M_{2\times 4}$*. What is the dimension of* $M_{2\times 4}$*? In general, what is the dimension of* $M_{m\times n}$*?*

Example 4.7.2. *Consider the subset* T *of* $M_{2\times 2}$ *of matrices whose diagonal elements sum to zero. That is,*

$$T = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \,\middle|\, a + d = 0 \right\}.$$

Show that T *is a subspace of* $M_{2\times 2}$.

We need to establish that the set T is nonempty and closed under vector addition and scalar multiplication. We can immediately see that the zero matrix,

$$O_{2\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

is an element of T, so T is clearly nonempty. To show that T is closed under vector addition, we need to show that if we take any two elements of T, say A and B, then their sum $A+B$ would also satisfy the condition necessary to belong to T. To this end, suppose

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

are in T. This means that

$$a_{11} + a_{22} = 0 \quad \text{and} \quad b_{11} + b_{22} = 0.$$

The sum $A+B$ is

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix}.$$

When we sum the diagonal entries, we can rearrange the terms in the sum to find that

$$(a_{11} + b_{11}) + (a_{22} + b_{22}) = (a_{11} + a_{22}) + (b_{11} + b_{22}) = 0 + 0 = 0.$$

So the sum of the diagonal entries of $A+B$ is zero, meaning that $A+B$ is an element of T. T is therefore closed under vector addition.

Next we show that T is closed under scalar multiplication. Letting c be any scalar, we have

$$cA = \begin{bmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{bmatrix}.$$

Summing the diagonal entries of cA, we get

$$ca_{11} + ca_{22} = c(a_{11} + a_{22}) = c(0) = 0.$$

This shows that T is indeed closed under scalar multiplication. We can conclude that T is a subspace of $M_{2\times 2}$.

Remark 4.7.1. *There is a special name for the sum of the diagonal entries of a matrix. It's called the* trace *of the matrix, and elements of the set T in Example 4.7.2 are called* trace-free *matrices.*

Example 4.7.3. *The subspace, T, of trace-free matrices in $M_{2\times 2}$ described in Example 4.7.2 can be expressed in terms of a span. An example of a spanning set is the set*

$$S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}.$$

If A is a trace free matrix, then it's diagonal entries must be additive inverses (the same number with opposite signs). The off diagonal entries can be any real numbers. We can express any such matrix as a linear combination of the elements of S.

$$A = \begin{bmatrix} a & b \\ c & -a \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Exercise 4.7.7. *In Example 4.7.3 we showed that the set of matrices*

$$S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$$

spans the subspace of trace–free matrices T. In other words, $\text{Span}(S) = T$.

1. *Show that S is linearly independent and is thus a basis for T.*

2. *What is the dimension of T?*

3. *What is the coordinate vector of the matrix*

$$A = \begin{bmatrix} 4 & 1 \\ 12 & -4 \end{bmatrix}$$

with respect to the basis S? (In other words, what is $[A]_S$?)

Exercise 4.7.8. *Let Z_s be the subset of $M_{2\times 2}$ whose entries sum to zero. That is,*

$$Z_s = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \,\middle|\, a + b + c + d = 0 \right\}.$$

1. *Show that* Z_s *is a subspace of* $M_{2\times 2}$.

2. *Find a basis,* $\mathcal{B}$*, for* Z_s.

3. *What is the dimension of* Z_s*?*

4. *What is the coordinate vector of the matrix*

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

with respect to the basis $\mathcal{B}$*?*

Exercise 4.7.9. *Let* N_s *be the subset of* $M_{2\times 2}$ *whose entries sum to one. That is,*

$$N_s = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \;\middle|\; a+b+c+d = 1. \right\}.$$

Show that N_s *is not a subspace of* $M_{2\times 2}$.

Exercise 4.7.10. *Let* D *be the subset of* $M_{2\times 2}$ *that consists of all diagonal matrices. That is*

$$D = \left\{ \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \;\middle|\; a, d \in R \right\}.$$

1. *Show that* D *is a subspace of* $M_{2\times 2}$.

2. *Find a basis,* $\mathcal{B}$*, for* D.

3. *What is the dimension of* D*?*

4. *What is the coordinate vector of the matrix*

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

with respect to the basis $\mathcal{B}$*?*

4.7.3 The Vector Space R^∞

You might guess what the symbol R^∞ stands for. When n is a positive integer, R^n denotes the set of ordered n–tuples of real numbers. Thus R^∞ denotes the set of all infinite sequences of real numbers. The elements of R^∞ have the form

$$\vec{a} = \langle a_1, a_2, a_{3,} \ldots \rangle .$$

Note that we are choosing to retain the arrow notation that we used in R^n. However, we are choosing to use letters from the beginning of the alphabet, such as a and b, rather than letters from the end of the alphabet, such as x and y, to name the vectors in R^∞ because these are the letters that are typically used when working with infinite sequences of real numbers in calculus courses.

We define vector addition and scalar multiplication in R^∞ just as in R^n: If $\vec{a} = \langle a_1, a_2, a_{3,} \ldots \rangle$ and $\vec{b} = \langle b_1, b_2, b_{3,} \ldots \rangle$ are elements of R^∞ and c is a scalar, then

$$\vec{a} + \vec{b} = \langle a_1 + b_1, a_2 + b_2, a_3 + b_3, \ldots \rangle$$

and

$$c\vec{a} = \langle ca_1, ca_2, ca_{3,} \ldots \rangle .$$

It is straightforward to check that R^∞ with the operations defined above satisfies all of the axioms of Definition 4.6.1 and is thus a real vector space. The zero vector of R^∞ is

$$\vec{0} = \langle 0, 0, 0, \ldots \rangle$$

and the additive inverse of $\vec{a} = \langle a_1, a_2, a_{3,} \ldots \rangle$ is

$$-\vec{a} = \langle -a_1, -a_2, -a_{3,} \ldots \rangle .$$

R^∞ is our first example of a vector space that is infinite dimensional. It has no basis that consists of a finite number of vectors. To get a feel for why this is so, let's just look at the two element set $S = \left\{\vec{a}, \vec{b}\right\}$ where $\vec{a}$ is the infinite sequence with $a_n = n$ $(1 \leq n < \infty)$ and $\vec{b}$ is the infinite sequence with $b_n = (-1)^n$ $(1 \leq n < \infty)$. Thus

$$\begin{aligned} \vec{a} &= \langle 1, 2, 3, 4, 5, \ldots \rangle \\ \vec{b} &= \langle -1, 1, -1, 1, -1, \ldots \rangle . \end{aligned}$$

Any linear combination of $\vec{a}$ and $\vec{b}$ has the form

$$\begin{aligned} s\vec{a} + t\vec{b} &= \langle s, 2s, 3s, 4s, 5s, \ldots \rangle + \langle -t, t, -t, t, -t, \ldots \rangle \\ &= \langle s - t, 2s + t, 3s - t, 4s + t, 5s - t, \ldots \rangle . \end{aligned}$$

For example,

$$\begin{aligned} 2\vec{a} + 3\vec{b} &= \langle 2 - 3, 2\,(2) + 3, 3\,(2) - 3, 4\,(2) + 3, 5\,(2) - 3, \ldots \rangle \\ &= \langle -1, 7, 3, 11, 7, \ldots \rangle . \end{aligned}$$

We have just constructed an example of a vector that is in Span $\left\{\vec{a}, \vec{b}\right\}$. However, it is also easy to construct many examples of vectors that are *not* in Span $\left\{\vec{a}, \vec{b}\right\}$. An example of such a vector is $\vec{c} = \langle -1, 0, 0, 0, 0, \ldots \rangle$. (This vector has -1 as its first component and all other components are 0.) To see why the vector $\vec{c}$ is not in Span $\left\{\vec{a}, \vec{b}\right\}$, let's look at the equation $s\vec{a} + t\vec{b} = \vec{c}$. Written out in long form, this equation is

$$\langle s - t, 2s + t, 3s - t, 4s + t, 5s - t, \ldots \rangle = \langle -1, 0, 0, 0, 0, \ldots \rangle .$$

For the vectors in the above equation to be equal, we must have

$$\begin{aligned} s - t &= -1 \\ 2s + t &= 0 \\ 3s - t &= 0 \\ &\vdots \\ &\text{etc.} \end{aligned}$$

The first equation above tells us that we must have $t = s + 1$ and when we substitute this into the second equation, we obtain

$$2s + (s + 1) = 0$$

which gives $s = -1/3$. Substitution back into $t = s + 1$ gives $t = 2/3$. However, when we substitute into the third equation, we obtain

$$3s - t = 3\left(-\frac{1}{3}\right) - \frac{2}{3} = -\frac{5}{3} \neq 0.$$

Thus there are no scalars s and t for which $s\vec{a} + t\vec{b} = \vec{c}$ is true. This means that $\vec{c} \notin \text{Span}\left\{\vec{a}, \vec{b}\right\}$, which means that $\text{Span}\left\{\vec{a}, \vec{b}\right\} \neq R^\infty$ and thus $\left\{\vec{a}, \vec{b}\right\}$ is not a basis for R^∞.

We have shown a specific example of a set of two vectors in R^∞ that is *not* a basis for R^∞. This certainly does not prove that R^∞ is infinite dimensional. It only proves that this particular set of two vectors is not a basis for R^∞. However, it is true that there is no finite set of vectors in R^∞ that spans R^∞. With some thought, you can probably come up with a proof of this fact. As a suggestion of how you might come up with a proof, start by trying to prove that no set of two vectors can span R^∞.

Exercise 4.7.11. *Let* $S = \left\{\vec{a}, \vec{b}\right\}$ *where* $a_n = 3$ *for all* n *and* $b_n = (-1)^n n^2$ *for all* n*. Thus*

$$\vec{a} = \langle 3, 3, 3, 3, 3, \ldots \rangle$$
$$\vec{b} = \langle -1, 4, -9, 16, -25, \ldots \rangle .$$

1. *Explain why the vector* $\vec{0} = \langle 0, 0, 0, 0, 0, \ldots \rangle$ *is in* $\text{Span}\left\{\vec{a}, \vec{b}\right\}$.

2. *Come up an example of a non–zero vector* $\vec{c}$ *that is in* $\text{Span}\left\{\vec{a}, \vec{b}\right\}$.

3. *Prove that* $\text{Span}\left\{\vec{a}, \vec{b}\right\} \neq R^\infty$.

Exercise 4.7.12. *Prove that the set of vectors* $S = \left\{\vec{a}, \vec{b}\right\}$ *given in Exercise 4.7.11 is linearly independent.*

If you have taken two semesters of calculus, you have probably studied infinite sequences and the concept of convergence of an infinite sequence. An infinite sequence, $\vec{a} = \langle a_1, a_2, a_3, \ldots \rangle$ is said to **converge** if there is a real number A such that

$$\lim_{n\to\infty} a_n = A.$$

The number A is called the **limit** of the sequence $\vec{a}$. As an example, if

$$a_n = \frac{2n}{3n+5},$$

then

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{2n}{3n+5} = \frac{2}{3}.$$

Thus $\vec{a}$ converges and has limit $A = 2/3$. An infinite sequence is said to **diverge** if it does not converge.

Two important basic facts (from calculus) about convergent sequences are that the sum of two convergent sequences is convergent and that any scalar multiple of a convergent sequence is convergent. In other words, if the sequences $\vec{a}$ and $\vec{b}$ both converge, then the sequence $\vec{a}+\vec{b}$ also converges and if the sequence $\vec{a}$ converges and c is any scalar, then the sequence $c\vec{a}$ also converges. To be more specific, if $\vec{a}$ has limit A and $\vec{b}$ has limit B, then $\vec{a}+\vec{b}$ has limit $A+B$ and $c\vec{a}$ has limit cA.

What we have just said in the above paragraph, when stated from the point of view of linear algebra, is that the set of all *convergent* sequences,

$$C = \{\vec{a} \in R^{\infty} \mid \vec{a} \text{ converges}\}$$

(which is a non–empty subset of R^{∞}) is closed under vector addition and also closed under scalar multiplication. Thus C is a subspace of R^{∞}! The subspace C is also infinite dimensional. (We will not prove this but perhaps you can come up with a proof if you think about it.)

We have now seen our first example of an infinite dimensional vector space, R^{∞}, and infinite dimensional subspace, C, of R^{∞}. It is important to point out that any infinite dimensional vector space also has finite dimensional subspaces. They are easy to construct. If we take any infinite–dimensional vector space V and take any non–zero vector $\vec{a} \in V$, then $\text{Span}\,\{\vec{a}\}$ is a subspace of dimension 1. As a specific example in R^{∞}, consider $\text{Span}\,\{\vec{a}\}$ where $\vec{a} = \langle 1, 2, 3, 4, 5, \ldots \rangle$. $\text{Span}\,\{\vec{a}\}$ is the set of all scalar multiples of $\vec{a}$, and $\{\vec{a}\}$ is a basis for $\text{Span}\,\{\vec{a}\}$. Since this basis contains only one vector, then $\dim(\text{Span}\,\{\vec{a}\}) = 1$.

Exercise 4.7.13. *1. Let* $\vec{a} = \langle 1, 2, 3, 4, 5, \ldots \rangle$ *and* $\vec{b} = \langle 2, 3, 4, 5, 6, \ldots \rangle$. *What is the dimension of the subspace* $S = \text{Span}\left\{\vec{a}, \vec{b}\right\}$*? Explain.*

2. Let $\vec{a} = \langle 1, 2, 3, 4, 5, \ldots \rangle$ *and* $\vec{b} = \langle 3, 6, 9, 12, 15, \ldots \rangle$. *What is the dimension of* $S = \text{Span}\left\{\vec{a}, \vec{b}\right\}$*? Explain.*

3. Let $\vec{e}_1 = \langle 1, 0, 0, 0, 0, \ldots \rangle$, $\vec{e}_2 = \langle 0, 1, 0, 0, 0, \ldots \rangle$ *and* $\vec{e}_3 = \langle 0, 0, 1, 0, 0, \ldots \rangle$. *(Thus* $\vec{e}_1$ *has* 1 *as its first entry and all other entries are* 0*, etc.) What is the dimension of* $\text{Span}\,\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$*? Explain.*

4.7.4 Vector Spaces of Functions

In your study of calculus (and courses prior to calculus), you have become familiar with the concept of "function". You are probably familiar with functions such as $f(x) = 4x^2 - 3x + 8$, which is a polynomial function, $f(x) = 3e^x$, which is an exponential function, $f(x) = 4\cos(x)$, which is a trigonometric function, and many others. We are going to consider vector spaces whose vectors are functions. In order to do that, we first need to recall some basic ideas about the function concept. We will restrict our attention to real–valued functions of a real variable. These are functions that have formulas that look like $y = f(x)$ where both x and y are allowed to be real numbers. In other words, both the input of the function, x, and the output of the function, y, are real numbers.

First, let us recall what we mean by the *domain* of a function, f. The domain of f is the subset of R from which the allowable inputs of f come. We need to specify the domain of a function as part of the definition of the function. Suppose that we specify the domain to be some set D, where D is a subset of R. Then the notation

$$f : D \to R$$

is used to say that f is a function whose domain is D and whose outputs are real numbers. As a specific example, suppose that we say that

$$f : R \to R$$

is the function defined by the formula

$$f(x) = 4x^2 - 3x + 8.$$

In writing $f : R \to R$, we are stating that the domain of the function is R and that the outputs of the function are also real numbers.

An important issue that we need to keep in mind when we are studying vector spaces whose vectors are functions is to carefully consider what we mean when we say that two functions are *equal* to each other. We way that two functions, f and g, are equal to each other if f and g both have the same domain, D, *and* $f(x) = g(x)$ for all $x \in D$.

For example, consider the functions $f : R \to R$ and $g : R \to R$ defined by the formulas

$$\begin{aligned} f(x) &= \sin^2(x) + \cos^2(x) \\ g(x) &= 1. \end{aligned}$$

These functions are equal to each other! Why? Because they both have the same domain (R) and we know from trigonometry that if x is any real number, then $\sin^2(x) + \cos^2(x) = 1$, and thus $f(x) = g(x)$ for all x in R. We can write $f = g$.

As another example, consider the functions $f : R \to R$ and $g : [-1, 1] \to R$ defined by the formulas

$$f(x) = x^2$$
$$g(x) = x^2.$$

These functions are *not* equal to each other, because they have different domains. In this example, it is not correct to write $f = g$.

As one more example (for those who have taken Calculus II), the function $f : (-1, 1) \to R$ defined by the sum of the power series

$$f(x) = \sum_{n=1}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots$$

is equal to the function $g : (-1, 1) \to R$ defined by

$$g(x) = \frac{1}{1-x}.$$

You may recall (from calculus) that the type of infinite series defined by the formula for f given above is called a *geometric series*. It converges to the sum $1/(1-x)$ as long as we insist that x is chosen from the open interval $(-1, 1)$. If x is not in this interval, then the series diverges and the formula given for f makes no sense. That is why we have designated the domain of f to be $(-1, 1)$. When we choose $x \in (-1, 1)$, it is true that $f(x) = g(x)$. Since we designated the domain of g to also be $(-1, 1)$, we can write $f = g$. Note that the formula that defines g, which is $1/(1-x)$, is defined for all real number values of x except $x = 1$, but writing $f = g$ only makes sense if we restrict the domain to be $(-1, 1)$ because the formula for f does not make sense otherwise.

Exercise 4.7.14. *In each part below, a pair of functions, f and g, are given. Determine whether or not f and g are equal to each other.*

1. *$f : (0, \infty) \to R$ and $g : (-\infty, 0) \to R$ defined by the formulas*

$$f(x) = x^2$$
$$g(x) = x^2.$$

2. *$f : R \to R$ and $g : R \to R$ defined by the formulas*

$$f(x) = (x+3)^2$$
$$g(x) = x^2 + 6x + 9.$$

3. *$f : R \to R$ and $g : R \to R$ defined by the formulas*

$$f(x) = (x+1)^2$$
$$g(x) = x^2 + 1.$$

4. *$f : R \to R$ and $g : R \to R$ defined by the formulas*

$$f(x) = \sin(2x)$$
$$g(x) = 2\sin(x)\cos(x).$$

5. *$f : (0,\infty) \to R$ and $g : (0,\infty) \to R$ defined by the formulas*

$$f(x) = \frac{1}{x} - \frac{1}{1+x}$$
$$g(x) = \frac{1}{x^2+x}.$$

Another thing we need to address before giving examples of vector spaces whose vectors are functions is how we add two functions together and how we multiply a function by a scalar. This is necessary because when we define a vector space, we need to define the addition and scalar multiplication operations on that vector space.

For two functions $f : D \to R$ and $g : D \to R$, we define the **sum** $f + g$ to be the function with domain D defined by the formula

$$(f+g)(x) = f(x) + g(x). \tag{4.26}$$

If $f : D \to R$ and c is a scalar (a real number), then we define the **scalar multiple** cf to be the function with domain D defined by the formula

$$(cf)(x) = cf(x). \tag{4.27}$$

As specific examples, suppose that $f : R \to R$ and $g : R \to R$ are the functions defined by

$$f(x) = 3x^2 - 4x + 4$$
$$g(x) = -4x^2 + 4x + 4.$$

Then $f+g$ is the function with domain R defined by the formula

$$(f+g)(x) = f(x) + g(x) = \left(3x^2 - 4x + 4\right) + \left(-4x^2 + 4x + 4\right) = -x^2 + 8$$

and $2f$ is the function with domain R defined by the formula

$$(2f)(x) = 2(f(x)) = 2\left(3x^2 - 4x + 4\right) = 6x^2 - 8x + 8.$$

Other things that we need to have on hand if we are to define a vector space whose vectors are real–valued functions are a zero vector and additive inverses. As you might guess, the zero vector is the function that is identically equal to 0 for all x in the domain D. Since we are using lower case letters to denote functions, we will use the letter z to denote the zero function. Thus $z : D \to R$ is the function defined by the formula

$$z(x) = 0. \tag{4.28}$$

It is easy to see that this function serves as an additive identity for vector addition. If f is any function with domain D, then

$$f + z = f.$$

For any function f with domain D, the additive inverse of f is the function (with domain D) denoted by $-f$ and defined by

$$(-f)(x) = -(f(x)). \tag{4.29}$$

Having made this definition, we see that for any function f we have

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

As a specific example, the additive inverse of $f(x) = 3x^2 - 4x + 4$ is

$$(-f)(x) = -\left(3x^2 - 4x + 4\right) = -3x^2 + 4x - 4.$$

We are now prepared to provide examples of vector spaces whose vectors are functions. Vector spaces whose vectors are functions are often referred to as **function spaces**.

4.7.4.1 The Function Spaces $F(D)$

Suppose that D is some domain (meaning that D is some specified non–empty subset of R). We define the vector space $F(D)$ to be the vector space whose vectors are the set of **all** real valued functions with domain D and whose operations of vector addition and scalar multiplication are as defined in (4.26) and (4.27). Thus

$$F(D) = \{f \mid f : D \to R\}.$$

The zero vector and additive inverses in $F(D)$ are as defined in (4.28) and (4.29).

As an example, $F(R)$ is the vector space of all real–valued functions that have domain R. It contains functions defined by formulas such as $f(x) = x^2$, $f(x) = -3e^x + \sin(4x) - 12$, and $f(x) = 47$. It is obviously a big vector space! It is infinite dimensional. Note that $F(R)$ does not contain the function $f(x) = 1/x$ because this function is not defined at $x = 0$. (Thus the domain of this function is not R.) Also, $F(R)$ does not contain the function $f(x) = \ln(x)$. (Can you explain why?) As another specific example, $F((0,\infty))$ is the set of all real–valued functions whose domain is the interval $(0,\infty)$. This vector space does contain the functions $f(x) = 1/x$ and $f(x) = \ln(x)$. (Can you explain why?)

Exercise 4.7.15. *In each part below, two functions, f and g, in $F(R)$ are given. Scalars, c and d, are also given. Compute the linear combination $cf + dg$. The first one is done as an example.*

1. $f(x) = -2x^2 - 3x + 1,\ g(x) = -x^2 - 2x - 2,\ c = 3,\ d = -2.$

   Solution: $3f - 2g$ *is the function with domain R defined by the formula*

$$\begin{aligned}(3f - 2g)(x) &= (3f)(x) - (2g)(x)\\ &= 3f(x) - 2g(x)\\ &= 3\left(-2x^2 - 3x + 1\right) - 2\left(-x^2 - 2x - 2\right)\\ &= -4x^2 - 5x + 7.\end{aligned}$$

2. $f(x) = x^2 + x - 7,\ g(x) = e^x + 2,\ c = -2,\ d = 1$

3. $f(x) = \sin^2(x),\ g(x) = \cos^2(x),\ c = 1,\ d = 1$

4. $f(x) = -x^2 + 3x - 2,\ g(x) = x^2 - 3x + 2,\ c = 1,\ d = 1$

5. $f(x) = -x^2 + 3x - 2,\ g(x) = 0,\ c = 1,\ d = 4$

6. $f(x) = -3x^2 - x - 3,\ g(x) = x^2 - 3x + 2,\ c = 0,\ d = 1$

Exercise 4.7.16. *Which of the following functions are in $F(R)$? If the function is not in $F(R)$, explain why not.*

1. $f(x) = e^x$

2. $f(x) = -12x^3 - 7x + 1$

3. $f(x) = \frac{1}{x^2+1}$

4. $f(x) = \frac{1}{x^2-1}$

5. $f(x) = \sqrt{x}$

Example 4.7.4. *Let $S = \{f, g\}$ where f and g are the functions in $F(R)$ defined by*

$$\begin{aligned} f(x) &= \sin^2(x) \\ g(x) &= \cos^2(x). \end{aligned}$$

Explain why the function h defined by $h(x) = 1$ is in $\mathrm{Span}(S)$.

Explanation: *Because of the trigonometric identity $\sin^2(x) + \cos^2(x) = 1$, which holds for all real numbers x, we have $f(x) + g(x) = h(x)$ for all real numbers x and thus we can write*

$$h = f + g$$

or

$$h = (1) f + (1) g$$

which shows that h is a linear combination of the functions in S and thus $h \in \mathrm{Span}(S)$.

Remark 4.7.2. *Referring to Example 4.7.4, it is somewhat cumbersome (as far as having to write so much) when we define a set of functions by saying $S = \{f, g\}$ and then giving the formulas such as $f(x) = \sin^2(x)$ and $g(x) = \cos^2(x)$ for f and g. It is easier to just start out by saying* *that* $S = \left\{\sin^2(x), \cos^2(x)\right\}$*, with the understanding that we are taking a shortcut in writing to save space. Then to show that the function* $h(x) = 1$ *is in* $\text{Span}(S)$*, we can just write*

$$1 = (1)\sin^2(x) + (1)\cos^2(x) \quad \text{for all} \quad x \in R.$$

We will write the next example (and future examples) using this convention, and you can use it in doing the exercises.

Example 4.7.5. *In the vector space* $F(R)$*, let* $S = \{1, x, x^2, x^3\}$ *and let* h *be the function* $h(x) = 2 + x - x^2$*. Explain why* $h \in \text{Span}(S)$*.*

Explanation: *We see that*

$$h(x) = (2)\,1 + (1)\,x + (-1)\,x^2 + (0)\,x^3 \text{ for all } x \in R$$

and thus h *is a linear combination of the functions in* S*. Hence* $h \in \text{Span}(S)$*.*

4.7.4.2 The Function Spaces $C^n(I)$

Some vector spaces that appear in applications, especially in the study of differential equations, are sets of functions with specified continuity or differentiability. For some given interval I, the set $C^0(I)$ is the set of all real–valued functions that are continuous on the domain I. For example, $C^0(R)$ is the set of all functions that are continuous on the whole real line, and $C^0([0,1])$ is the set of all functions that are continuous on the interval $[0,1]$. Vector addition and scalar multiplication in $C^0(I)$ are as defined in (4.26) and (4.27). The zero vector and additive inverses in $C^0(I)$ are as defined in (4.28) and (4.29). Because the sum of continuous functions is continuous and a scalar multiple of a continuous function is continuous (facts from calculus), the set $C^0(I)$ is closed under both of these operations. Thus $C^0(I)$ is a subspace of $F(I)$.

The set $C^1(I)$ is the set of all real valued functions that are at least one time continuously differentiable on the interval I. (To say that a function f is continuously differentiable on I means that f is differentiable on I and that the derivative of f is also a continuous function on I.) You might remember from calculus that differentiability implies continuity. Hence $C^1(I)$ is a subset of $C^0(I)$. Furthermore, the sum of two continuously differentiable functions is continuously differentiable and a scalar multiple of a continuously differentiable function is continuously differentiable, and that tells us that $C^1(I)$ is in fact a subspace of $C^0(I)$.

Similar notation is used to denote the set of functions having at least a specific number of continuous derivatives: Elements of $C^2(I)$ are functions that are at least twice continuously differentiable; elements of $C^3(I)$ are at least three times continuously differentiable, and so forth. The elements of $C^\infty(I)$ are functions that have continuous derivatives of all orders. Many of the functions encountered in calculus, such as e^x, $\cos(x)$, $\tan^{-1}(x)$, and polynomials are in the vector space $C^\infty(R)$.

Exercise 4.7.17. *Let F_0 be the subset of $C^0(R)$ of functions that take the value of zero at zero. That is,*

$$F_0 = \left\{ f \in C^0(R) \mid f(0) = 0 \right\}.$$

Determine whether F_0 is a subspace of $C^0(R)$. That is, either show that F_0 is a subspace of $C^0(R)$, or demonstrate that F_0 is not closed under vector addition or scalar multiplication.

4.7.4.3 Function Spaces of Polynomials

If V is any vector space, either finite dimensional or infinite dimensional, we can always construct a finite dimensional subspace of V by just choosing some set of vectors out of V and forming the span of this set. If V is a vector space and $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is a finite set of k vectors in V, then $\text{Span}\,(S)$ is a subspace of V with $\dim\left(\text{Span}\,(S)\right) \leq k$. If the set S is linearly independent, then $\dim\left(\text{Span}\,(S)\right) = k$. Otherwise $\dim\left(\text{Span}\,(S)\right) < k$.

The function spaces $\mathbb{P}_n$ are the subspaces of $F\,(R)$ defined by

$$\begin{aligned}
\mathbb{P}_0 &= \text{Span}\,\{1\} \\
\mathbb{P}_1 &= \text{Span}\,\{1, x\} \\
\mathbb{P}_2 &= \text{Span}\left\{1, x, x^2\right\} \\
\mathbb{P}_3 &= \text{Span}\left\{1, x, x^2, x^3\right\} \\
&\text{etc.}
\end{aligned}$$

So, for example, $\mathbb{P}_2$ consists of all functions that have domain R and are defined by formulas of the form

$$p\,(x) = a_0 + a_1 x + a_2 x^2,$$

where a_0, a_1, and a_2 can be any scalars. This is the set of all polynomial functions that have degree 2 or less. (We need to say degree 2 **or less** because if $a_2 = 0$ then there is no x^2 term in the polynomial and that means that the degree of the polynomial is less than 2.) Likewise $\mathbb{P}_3$ is the space of all polynomial functions that have degree 3 or less. The functions in $\mathbb{P}_3$ have the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3.$$

Exercise 4.7.18. *Explain why $\mathbb{P}_2$ is a subspace of $\mathbb{P}_3$.*

Exercise 4.7.19. *Suppose we had decided to define $\mathbb{P}_2$ to be the set of all polynomial functions whose degree is **exactly** 2. This would mean that we were defining $\mathbb{P}_2$ to be the set of all polynomial functions of the form*

$$p(x) = a_0 + a_1 x + a_2 x^2$$

*where a_0 and a_1 can be any scalars and a_2 can be any scalar except 0. Explain why this choice of definition would result in $\mathbb{P}_2$ **not** being a subspace of $F(R)$.*

Exercise 4.7.20. *Consider the vector space $\mathbb{P}_4$.*

1. *Determine whether the functions defined by the given formulas are vectors in $\mathbb{P}_4$.*

   (a) $p(x) = 2 + 3x - x^2 + 2x^3 + 4x^4$

   (b) $q(x) = 2 + 3x^2 - 9x^3 + 2x^4$

   (c) $f(x) = -12 + x + 5x^2 - 6x^3$

   (d) $r(x) = 21x^3 - 4x^5$

2. *Let f and g be the functions defined by $f(x) = 2x + x^3 - 14x^4$ and $g(x) = -3 + 4x^2 - 5x^3 + 10x^4$.*

   (a) *evaluate $2f$*

   (b) *evaluate $3g$*

   (c) *evaluate $f - g$*

   (d) *What is the additive inverse of g?*

Exercise 4.7.21. *Let $\mathcal{P}_{2,1}$ denote the set of all polynomials, $p(x) = p_0 + p_1 x + p_2 x^2$ of degree at most 2 with real coefficients that satisfy $p(1) = 0$.*

1. *Determine which of the following are elements of* $\mathcal{P}_{2,1}$.

   (a) $g(x) = 2 - 3x - x^2$
   (b) $f(x) = 2 - 3x + x^2$
   (c) $q(x) = 4x^2 + 2x - 6$

2. *Show that* $\mathcal{P}_{2,1}$ *is closed with respect to vector addition and scalar multiplication.*

3. *Verify that every element of* $\mathcal{P}_{2,1}$ *can be written in the form* $p(x) = p_1(x-1) + p_2(x^2-1)$. *Note that we can say that* $\mathcal{P}_{2,1} = \mathrm{Span}\{x-1, x^2-1\}$.

Example 4.7.6. *Consider the subset* $\mathcal{P}_{2,1}$ *of* $\mathbb{P}_2$ *containing polynomials* $p(x) = p_0 + p_1 x + p_2 x^2$ *with the property that* $p(1) = 0$ *from Exercise 4.7.21 above. This set is nonempty, in particular it contains the zero vector* $z(x) = 0 + 0x + 0x^2$. *In Exercise 4.7.21, you established that* $\mathcal{P}_{2,1}$ *is closed with respect to vector addition and scalar multiplication. Hence* $\mathcal{P}_{2,1}$ *is a subspace of* $\mathbb{P}_2$.

We have defined $\mathbb{P}_2 = \mathrm{Span}\{1, x, x^2\}$. Thus $\mathbb{P}_2$ is the span of a set of three functions. It is thus natural to ask whether or not the dimension of $\mathbb{P}_2$ is 3. Indeed it is true that $\dim(\mathbb{P}_2) = 3$, but in order to verify this we need to show that the set of vectors $S = \{1, x, x^2\}$ is linearly independent. We do this in the following example.

Example 4.7.7. *Let* $S = \{p_0, p_1, p_2\}$ *be the set of functions defined by* $p_0(x) = 1$, $p_1(x) = x$ *and* $p_2(x) = x^2$. *Verify that the set* S *is a basis for* $\mathbb{P}_2$.

Solution: *We already know that* $\mathrm{Span}(S) = \mathbb{P}_2$, *because this is how we have defined* $\mathbb{P}_2$. *Thus we only need to show that* S *is linearly independent. This means we need to show that the equation*

$$c_0 p_0 + c_1 p_1 + c_2 p_2 = z$$

has only the trivial solution $c_0 = c_1 = c_2 = 0$. *The above equation is equivalent to*

$$c_0 p_0(x) + c_1 p_1(x) + c_2 p_2(x) = z(x) \text{ for all } x \in R.$$

Thus the equation that we need to study (and show has only the trivial solution) is

$$c_0(1) + c_1 x + c_2 x^2 = 0 \text{ for all } x \in R. \tag{4.30}$$

If the equation (4.30) is to be true for all $x \in R$, then that means we can choose any value of x that we like and the equation needs to be true for that value of x. If we choose $x = 0$, then we obtain

$$c_0(1) + c_1(0) + c_2(0)^2 = 0$$

which tells us that we must have $c_0 = 0$.

Knowing that $c_0 = 0$, we now have reduced the problem to studying the equation

$$c_1 x + c_2 x^2 = 0 \text{ for all } x \in R. \tag{4.31}$$

Since equation (4.31) must be true for all $x \in R$, then it must be true for $x = 1$. Plugging $x = 1$ into the above equation gives

$$c_1(1) + c_2(1)^2 = 0$$

and this tells us that we must have $c_1 + c_2 = 0$.

Since equation (4.31) must be true for $x = -1$, we obtain

$$c_1(-1) + c_2(-1)^2 = 0$$

and this tells us that we must have $-c_1 + c_2 = 0$.

We now have a system of two equations to solve:

$$\begin{aligned} c_1 + c_2 &= 0 \\ -c_1 + c_2 &= 0 \end{aligned}$$

and it is easily seen that the only solution of this system of equations is $c_1 = c_2 = 0$.

We have shown that the equation

$$c_0 p_0 + c_1 p_1 + c_2 p_2 = z$$

has only the trivial solution $c_0 = c_1 = c_2 = 0$, and thus we have shown that S is linearly independent. Therefore S is a basis for $\mathbb{P}_2$. Since S is a basis for $\mathbb{P}_2$ and S contains three vectors, then $\dim(\mathbb{P}_2) = 3$.

Example 4.7.8. *Consider the basis $S = \{p_0, p_1, p_2\}$ for $\mathbb{P}_2$ consisting of the vectors $p_0(x) = 1$, $p_1(x) = x$ and $p_2(x) = x^2$, in this order, from Example 4.7.7.*

1. *Identify the coordinate vector,* $[p]_S$*, for* $p(x) = 2 - 4x + 7x^2$*.*

2. *Identify the coordinate vector,* $[q]_S$*, for* $q(x) = -3 + 12x - 5x^2$*.*

3. *Evaluate the sum* $p(x) + q(x)$*, and find its coordinate vector ,* $[p + q]_S$*.*

4. *Compare the sum of the coordinate vectors from parts 1. and 2. with the coordinate vector found in part 3.*

5. *Evaluate* $3p(x)$*, and find its coordinate vector,* $[3p]_S$*. Compare this result to* 3 *times the coordinate vector for* p *that you found in part 1.*

Solutions:

1. *We can write* $p(x) = 2 - 4x + 7x^2 = 2p_0(x) - 4p_1(x) + 7p_2(x)$*. So the entries of the coordinate vector are these coefficients* 2*,* -4*, and* 7*, in that order. That is,*
$$[p]_S = \langle 2, -4, 7 \rangle.$$

2. *Following the same process,*
$$[q]_S = \langle -3, 12, -5 \rangle.$$

3. *If we add the vectors* p *and* q*, we get*
$$p(x) + q(x) = (2 - 3) + (-4 + 12)x + (7 - 5)x^2 = -1 + 8x + 2x^2.$$
The coordinate vector for the sum is
$$[p + q]_S = \langle -1, 8, 2 \rangle.$$

4. *Now we're asked to sum the coordinate vectors from* R^3 *that we found in parts 1. and 2.*
$$[p]_S + [q]_S = \langle 2, -4, 7 \rangle + \langle -3, 12, -5 \rangle = \langle -1, 8, 2 \rangle.$$
When we compare this to the coordinate vector we find in part 3., we see that they match.

5. *Scaling p by 3 gives $3p(x) = 6 - 12x + 21x^2$. So*

$$[3p]_S = \langle 6, -12, 21 \rangle.$$

If we scale the coordinate vector we found in part 1., we get

$$3[p]_S = 3\langle 2, -4, 7 \rangle = \langle 6, -12, 21 \rangle.$$

These also match, $[3p]_S = 3[p]_S$.

Exercise 4.7.22. *Show that the set of vectors $S = \{1, x, x^2, x^3\}$ is a basis for $\mathbb{P}_3 = \text{Span}\{1, x, x^2, x^3\}$ and hence that $\dim(\mathbb{P}_3) = 4$ (We already know that S spans $\mathbb{P}_3$, by definition).*

Hint: Follow the approach used in Example 4.7.7.

In Example 4.7.7 we showed that $S = \{1, x, x^2\}$ is a basis for $\mathbb{P}_2$ and hence $\dim(\mathbb{P}_2) = 3$. In Exercise 4.7.22 you were asked to show that $S = \{1, x, x^2, x^3\}$ is a basis for $\mathbb{P}_3$ and hence $\dim(\mathbb{P}_3) = 4$. There is clearly a pattern here. In general, the set $S = \{1, x, \ldots, x^n\}$ is a basis for $\mathbb{P}_n$. Since this basis for $\mathbb{P}_n$ contains exactly $n + 1$ vectors, then $\dim(\mathbb{P}_n) = n + 1$

Example 4.7.9. *Show that the subset $S = \{p, q, r\}$ of $\mathbb{P}_2$ is linearly independent, where*

$$p(x) = 1 - x + x^2, \quad q(x) = 2 - x, \quad \text{and} \quad r(x) = 3 + x^2.$$

Solution: *We need to show that*

$$c_1 p + c_2 q + c_3 r = z$$

has only the trivial solution $c_1 = c_2 = c_3 = 0$.

This means that we need to show that

$$c_1 p(x) + c_2 q(x) + c_3 r(x) = z(x) \text{ for all } x \in R$$

has only the trivial solution, which means that we need to show that

$$c_1 \left(1 - x + x^2\right) + c_2 (2 - x) + c_3 \left(3 + x^2\right) = 0 \text{ for all } x \in R$$

has only the trivial solution.

By rearranging the above equation (organizing according to powers of x), we obtain

$$(c_1 + 2c_2 + 3c_3)(1) + (-c_1 - c_2)x + (c_1 + c_3)x^2 = 0 \text{ for all } x \in R.$$

Now recall that we showed in Example 4.7.7 that the set of functions $\{1, x, x^2\}$ is a basis for $\mathbb{P}_2$.

This tells us that the all of the weights corresponding to 1, x, and x^2 in the above equation must be equal to 0. Therefore we must have

$$\begin{array}{rcrcrcl} c_1 & + & 2c_2 & + & 3c_3 & = & 0 \\ -c_1 & - & c_2 & & & = & 0 \\ c_1 & & & + & c_3 & = & 0 \end{array}.$$

Setting up the augmented matrix and performing row reduction,

$$\left[\begin{array}{rrr|r} 1 & 2 & 3 & 0 \\ -1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right] \xrightarrow{rref} \left[\begin{array}{rrr|r} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right].$$

We see that the system has unique solution $c_1 = c_2 = c_3 = 0$, the trivial solution. Hence $S = \{p, q, r\}$ is linearly independent.

Exercise 4.7.23. *Determine whether the indicated set is linearly independent or linearly dependent in the indicated vector space.*

1. $\{1 + x, 1 - x\}$ *in* $\mathbb{P}_1$.
2. $\{1 + x, 1 - x, 2 - 3x\}$ *in* $\mathbb{P}_1$.
3. $\{1 + 2x^2, -1 + x, -3 + 3x\}$ *in* $\mathbb{P}_2$
4. $\{-1 + 2x - x^2, 2 + x - 2x^3, 2 + 2x - 2x^3, 1 + x - x^2 + 2x^3\}$ *in* $\mathbb{P}_3$
5. $\{7\}$ *in* $\mathbb{P}_0$.

Example 4.7.10. *Find the coordinate vector of the function $p(x) = -5 + 6x + 6x^2$ with respect to the basis $S = \{1, x, x^2\}$.*

Solution: *The weights that are used to write p as a linear combination of the vectors in the ordered basis S are -5, 6, and 6 and thus the coordinate vector of p with respect to the ordered basis S is*

$$[p]_S = \langle -5, 6, 6 \rangle.$$

Example 4.7.11. *Consider the set of functions* $T = \{1, 1+x, 1+x+x^2\}$ *in* $\mathbb{P}_2$*. Show that* T *is a basis for* $\mathbb{P}_2$*. Then find* $[p]_T$ *where* p *is the function* $p(x) = -5 + 6x + 6x^2$*.*

Solution: *To show that* T *is linearly independent, we must show that the equation*

$$c_0 (1) + c_1 (1+x) + c_2 \left(1 + x + x^2\right) = 0 \;\; \text{for all } x \in R$$

has only the trivial solution.

First we gather like terms in the above equation to obtain

$$(c_0 + c_1 + c_2)(1) + (c_1 + c_2)\, x + c_2 x^2 = 0.$$

Since $S = \{1, x, x^2\}$ *is a basis for* $\mathbb{P}_2$*, all of the weights in the above equation must be* 0*. That is, we must have*

$$\begin{array}{ccccccc} c_1 & + & c_1 & + & c_2 & = & 0 \\ & & c_1 & + & c_2 & = & 0 \\ & & & & c_2 & = & 0 \end{array}.$$

It is easily seen that the only solution of the above system of equations is $c_0 = c_1 = c_2 = 0$*, and this proves that the set* T *is linearly independent.*

To show that $\text{Span}(T) = \mathbb{P}_2$*, we need to show that for any function* $p(x) = a_0 + a_1 x + a_2 x^2$ *in* $\mathbb{P}_2$*, there exist weights* c_0*,* c_1*, and* c_2 *such that*

$$c_0 (1) + c_1 (1+x) + c_2 \left(1 + x + x^2\right) = a_0 + a_1 x + a_2 x^2 \;\; \text{for all } x \in R.$$

Once again gathering like terms on the left hand side, we obtain

$$(c_0 + c_1 + c_2)(1) + (c_1 + c_2)\, x + c_2 x^2 = a_0 + a_1 x + a_2 x^2.$$

Since S *is a basis for* $\mathbb{P}_2$*, then we know that any element of* $\mathbb{P}_2$ *can be written uniquely as a linear combinations of the functions in* $S = \{1, x, x^2\}$*. This means that the corresponding weights on each side of the above equation must be equal to each other. That is we must have*

$$\begin{array}{ccccccc} c_1 & + & c_1 & + & c_2 & = & a_0 \\ & & c_1 & + & c_2 & = & a_1 \\ & & & & c_2 & = & a_2 \end{array}.$$

The above system of equations is easily solved by back substitution to obtain the unique solution

$$\begin{aligned} c_0 &= a_0 - a_1 \\ c_1 &= a_1 - a_2 \\ c_2 &= a_2. \end{aligned}$$

We have found that $p \in \operatorname{Span}(T)$ *for any* p *in* $\mathbb{P}_2$. *This, along with the fact that* T *is linearly independent, tells us that* T *is a basis for* $\mathbb{P}_2$. *We have also found that for any* p *in* $\mathbb{P}_2$ *that the coordinate vector of* p *with respect to* T *is*

$$[p]_T = \langle a_0 - a_1, a_1 - a_2, a_2 \rangle .$$

In particular (to answer the other question that was asked), if $p(x) = -5 + 6x + 6x^2$, *then*

$$[p]_T = \langle -5 - 6, 6 - 6, 6 \rangle = \langle -11, 0, 6 \rangle .$$

We can check this answer by noting that

$$-11(1) + 0(1 + x) + 6\left(1 + x + x^2\right) = -5 + 6x + 6x^2 \quad \text{for all } x \in R$$

is true.

Exercise 4.7.24. *Consider the set of functions* $S = \{3, 1 - x + x^2, 3 + 3x - 3x^2\}$ *in* $\mathbb{P}_2$. *Show that* S *is not linearly independent.*

4.7.4.4 Function Spaces Defined by Spans

Similar to what we did in defining the subspaces $\mathbb{P}_n$, we can take any finite set of functions in $F(R)$ and note that their span is a finite dimensional subspace of $F(R)$. This idea is relevant in the study of Differential Equations and other areas of Mathematics.

Example 4.7.12. *Let* S *be the set of functions* $S = \{\sin(x), \cos(x)\}$. *Then* $\operatorname{Span}(S)$ *is a subspace of* $F(R)$. *It is actually a subspace of* $C^\infty(R)$ *because the sine and cosine functions have continuous derivatives of all orders. Let us show that the set* S *is linearly independent.*

Solution: *We need to show that the equation*

$$c_1 \sin + c_2 \cos = z$$

has only the trivial solution $c_1 = c_2 = 0$.

This means that we need to show that

$$c_1 \sin(x) + c_2 \cos(x) = 0 \text{ for all } x \in R$$

has only the trivial solution.

Since we require that the above equation holds for all $x \in R$, *we can pick a specific value of* x *and plug it in. If we plug in* $x = 0$ *then we get*

$$c_1 \sin(0) + c_2 \cos(0) = 0$$

which simplifies to

$$c_1 (0) + c_2 (1) = 0$$

and thus we must have $c_2 = 0$. *The original equation is now reduced to*

$$c_1 \sin(x) = 0 \text{ for all } x \in R.$$

We can plug in $x = \pi/2$ *to obtain*

$$c_1 \sin\left(\frac{\pi}{2}\right) = 0,$$

and recalling that $\sin(\pi/2) = 1$ *we see that we must have* $c_1 = 0$. *We have shown that* S *is linearly independent and is thus a basis for* $\operatorname{Span}(S)$. *We see that* $\dim(\operatorname{Span}(S)) = 2$.

Those who have had a course in differential equations may recognize that $\operatorname{Span}(S)$ *is the solution set of the linear homogeneous differential equation*

$$y'' + y = 0.$$

What this means is that the functions in $\operatorname{Span}(S)$, *which are all functions of the form*

$$y(x) = c_1 \sin(x) + c_2 \cos(x),$$

are the complete set of functions that satisfy the above differential equation. In general, any linear homogeneous differential equation has a solution set that is of the form $\operatorname{Span}(S)$ *for some set of functions* S *in* $C^\infty(R)$.

Exercise 4.7.25. *Show that the set of functions* $S = \{e^x, e^{2x}\}$ *is linearly independent and thus* $\dim(\operatorname{Span}(S)) = 2$.

Exercise 4.7.26. *Show that the set of functions* $S = \{e^x, xe^x\}$ *is linearly independent and thus* $\dim(\operatorname{Span}(S)) = 2$.

Exercise 4.7.27. *Show that the set of functions* $S = \{1, \sin(x), \sin(2x)\}$ *is linearly independent and thus* $\dim(\operatorname{Span}(S)) = 3$.

4.7.4.5 The Vector Spaces R^n Viewed as Function Spaces

We have now seen several examples of finite dimensional and infinite dimensional subspaces of $F(R)$. We will now come full circle and once again consider the original vector spaces that we studied – the vector spaces R^n. The vector spaces R^n can actually be viewed as function spaces. To see why, suppose we define D_2 to be the set $D_2 = \{1, 2\}$. Then $F(D_2)$ is the set of all real valued functions with domain D_2. On the other hand, R^2 is the set of all ordered pairs of the form $\langle x_1, x_2\rangle$. We can think of $\langle x_1, x_2\rangle$ as being a function that assigns the subscript 1 to the real number x_1 and assigns the subscript 2 to the number x_2. If we call this function f, then f is a function in $F(D_2)$. As a specific example, the vector $\langle 5, -3\rangle$ in R^2, corresponds to the function in $F(D_2)$ defined by

$$\begin{aligned} f(1) &= 5 \\ f(2) &= -3. \end{aligned}$$

This function f is the function that assigns the subscript 1 to the number 5 and assigns the subscript 2 to the number -3.

In general, if we define the set D_n to be $D_n = \{1, 2, \ldots, n\}$, then R^n can be viewed as $F(D_n)$ with the vector $\langle x_1, x_2, \ldots, x_n\rangle$ in R^n being identified with the function

$$\begin{aligned} f(1) &= x_1 \\ f(2) &= x_2 \\ &\vdots \\ f(n) &= x_n \end{aligned}$$

in $F(D_n)$. Likewise, defining D_∞ to be the set of all positive integers, i.e., $D_\infty = \{1, 2, 3\ldots\}$, we can identity R^∞ with $F(D_\infty)$.

Viewing R^n and R^∞ as function spaces does not do much to help us understand R^n and R^∞ any better, but we feel that it is worth pointing out this connection. Understanding this connection may, if anything, be useful in helping you to understand the overall concept of function space better.

4.8 Working with Coordinate Vectors

In this section, we examine how we can use coordinate vectors as a tool that can be applied in working with any finite dimensional subspace of any vector space. The basic idea is that if we have an ordered basis, $\mathcal{B}$, for some subspace, S, of some vector space, V, then each vector, $\vec{x}$, in S can be written in a unique way as a linear combination of the vectors in $\mathcal{B}$. The vector of unique weights used in writing $\vec{x}$ as a linear combination of the vectors in $\mathcal{B}$ is the coordinate vector of $\vec{x}$ with respect to $\mathcal{B}$, denoted by $[\vec{x}]_{\mathcal{B}}$. If the ordered basis, $\mathcal{B}$, contains k vectors, then $\dim(S) = k$ and $[\vec{x}]_{\mathcal{B}}$ is a vector in R^k. If we wish to establish linear independence of a certain set of vectors, T, in S, then we can actually do this by establishing linear independence of the set of coordinate vectors of the vectors that make up T. The advantage of this is that it allows us to work in R^k and hence bring to bear all of the tools studied in Chapter 2, in particular the tools pertaining to augmented matrices and row reduction. The main ideas are illustrated in the theorems, examples and exercises that follow.

Lemma 4.8.1. *Suppose that S is a subspace of a vector space V and $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_k\}$ is an ordered basis of S. If $\vec{x}$ and $\vec{y}$ are any two vectors in S and c is any scalar then*

1. $[\vec{x} + \vec{y}]_{\mathcal{B}} = [\vec{x}]_{\mathcal{B}} + [\vec{y}]_{\mathcal{B}}$ *and*
2. $[c\vec{x}]_{\mathcal{B}} = c\,[\vec{x}]_{\mathcal{B}}$.

Proof. We will prove statement 1 and leave the proof of statement 2 as an exercise. Since $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_k\}$ is a basis for S, then every vector in S can be written in a unique way as a linear combination of the vectors in $\mathcal{B}$. Thus, there exists a unique set of scalars $c_1, c_2, \ldots, c_k$ and a unique set of scalars $d_1, d_2, \ldots, d_k$ such that

$$\vec{x} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k$$

and

$$\vec{y} = d_1\vec{u}_1 + d_2\vec{u}_2 + \cdots + d_k\vec{u}_k.$$

We thus have $[\vec{x}]_{\mathcal{B}} = \langle c_1, c_2, \ldots, c_k\rangle$ and $[\vec{y}]_{\mathcal{B}} = \langle d_1, d_2, \ldots, d_k\rangle$. Next we observe that

$$\vec{x} + \vec{y} = (c_1 + d_2)\,\vec{u}_1 + (c_2 + d_2)\,\vec{u}_2 + \cdots + (c_k + d_k)\,\vec{u}_k$$

and thus

$$\begin{aligned}[\vec{x} + \vec{y}]_{\mathcal{B}} &= \langle c_1 + d_1, c_2 + d_2, \ldots, c_k + d_k\rangle \\ &= \langle c_1, c_2, \ldots, c_k\rangle + \langle d_1, d_2, \ldots, d_k\rangle \\ &= [\vec{x}]_{\mathcal{B}} + [\vec{y}]_{\mathcal{B}}\end{aligned}$$

□

Exercise 4.8.1. *Prove statement 2 of Lemma 4.8.1.*

Lemma 4.8.2. *Suppose S is a subspace of a vector space V and $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_k\}$ is an ordered basis of S. Then $\vec{0}_V$ is the only vector in S that has coordinate vector $\vec{0}_k$. In other words, the following statement holds for all vectors $\vec{x} \in S$:*

$$[\vec{x}]_{\mathcal{B}} = \vec{0}_k \textit{ if and only if } \vec{x} = \vec{0}_V.$$

Proof. Suppose that $[\vec{x}]_{\mathcal{B}} = \vec{0}_k$. Then

$$\vec{x} = 0\vec{u}_1 + 0\vec{u}_2 + \cdots + 0\vec{u}_k.$$

By statement 3 of Theorem 4.6.1, $0\vec{u}_i = \vec{0}_V$ for all $i = 1, 2, \ldots k$ and thus

$$\vec{x} = \vec{0}_V + \vec{0}_V + \cdots + \vec{0}_V = \vec{0}_V.$$

Now suppose that $\vec{x} = \vec{0}_V$. Since $\mathcal{B}$ is a basis for S, then there exist weights $c_1, c_2, \ldots, c_k$ such that

$$\vec{x} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k$$

and thus

$$c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k = \vec{0}_V.$$

Since $\mathcal{B}$ is linearly independent, then $c_1 = c_2 = \cdots = c_k = 0$ and thus

$$\vec{x} = 0\vec{u}_1 + 0\vec{u}_2 + \cdots + 0\vec{u}_k.$$

This tells us that

$$[\vec{x}]_{\mathcal{B}} = \langle 0, 0, \ldots, 0 \rangle = \vec{0}_k.$$

□

Theorem 4.8.1. *Suppose S is a subspace of a vector space V and $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_k\}$ is an ordered basis of S. Let $T = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$ be any set of vectors in S, and let $C_T = \left\{[\vec{x}_1]_{\mathcal{B}}, [\vec{x}_2]_{\mathcal{B}}, \ldots, [\vec{x}_m]_{\mathcal{B}}\right\}$ be the set of vectors in R^k consisting of the coordinate vectors of the elements of T with respect to the basis $\mathcal{B}$. Then T is linearly independent in V if and only if C_T is linearly independent in R^k.*

Proof. Suppose that T is linearly independent in V. This means that the equation

$$c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_m\vec{x}_m = \vec{0}_V$$

has only the trivial solution. We want to prove that the set of vectors C_T is linearly independent in R^k. To do this, we need to show that the equation

$$c_1[\vec{x}_1]_{\mathcal{B}} + c_2[\vec{x}_2]_{\mathcal{B}} + \cdots + c_m[\vec{x}_m]_{\mathcal{B}} = \vec{0}_k \tag{4.32}$$

has only the trivial solution. Using Lemma 4.8.1 (both parts) we see that

$$c_1[\vec{x}_1]_{\mathcal{B}} + c_2[\vec{x}_2]_{\mathcal{B}} + \cdots + c_m[\vec{x}_m]_{\mathcal{B}} = [c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_m\vec{x}_m]_{\mathcal{B}}.$$

Thus equation (4.32) is equivalent to

$$[c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_m\vec{x}_m]_{\mathcal{B}} = \vec{0}_k.$$

By Lemma 4.8.2 it must be the case that

$$c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_m\vec{x}_m = \vec{0}_V$$

and we know that (by assumption) this equation has only the trivial solution. Therefore equation (4.32) has only the trivial solution, and this shows that the set of vectors C_T is linearly independent in R^k. The proof of the fact that if C_T is linearly independent in R^k then T is linearly independent in V is similar. □

Note what Theorem 4.8.1 is telling us. It is telling us that we can determine the linear dependence or independence of a set of vectors in some vector space V by examining the relationship between the coordinate vectors in R^k. This means that we can employ our tools from R^k, especially matrices! The following example provides an illustration of this.

Example 4.8.1. *We can use coordinate vectors with respect to the ordered basis $S = \{1, x, x^2\}$ to show that the subset $\{p, q, r\}$ of $\mathbb{P}_2$ is linearly independent, where*

$$p(x) = 1 - x + x^2, \quad q(x) = 2 - x, \quad \text{and} \quad r(x) = 3 + x^2.$$

First, we obtain the coordinate vectors with respect to S for each of the polynomials $p, q,$ and r. These are

$$[p]_S = \langle 1, -1, 1\rangle, \quad [q]_S = \langle 2, -1, 0\rangle, \quad \text{and} \quad [r]_S = \langle 3, 0, 1\rangle.$$

Theorem 4.8.1 tells us that the linear dependence or independence of these vectors in R^3 is equivalent to the linear dependence or independence of $\{p, q, r\}$ in $\mathbb{P}_2$. One way to investigate the linear independence in R^3 is to use a matrix. Using the coordinate vectors as column vectors, and using row reduction

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \xrightarrow{rref} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The matrix is row equivalent to the identity which shows that the columns—the coordinate vectors, are linearly independent. We conclude, as we did in Example 4.7.9, that the set $\{p, q, r\}$ is linearly independent in $\mathbb{P}_2$.

Exercise 4.8.2. *Let $\mathcal{B} = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ be the ordered basis of $M_{2\times 2}$ where*

$$E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Use coordinate vectors to determine whether the following collection of vectors is linearly dependent or linearly independent in $M_{2\times 2}$.

$$A_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 2 & -1 \\ 1 & -2 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 2 \\ 1 & 3 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 1 & 6 \\ 0 & 8 \end{bmatrix}.$$

Because linear dependence and independence can be determined by the relationship between coordinate vectors in some R^k, we can extend the results from Lemma 4.3.1 and Theorem 4.3.1 to general vector spaces. In particular, we have the following lemma.

Lemma 4.8.3. *Suppose S is a subspace of a vector space V and $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_k\}$ is an ordered basis of S. If $T = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ is any set of m vectors in S where $m > k$, then T is linearly dependent.*

This leads to the following theorem that confirms that all bases for a given subspace of a vector space must have the same number of elements.

Theorem 4.8.2. *Let $n \geq 2$ and $1 \leq k \leq n$. Suppose S is a subspace of a vector space V and $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_k\}$ is a basis of S. Then every basis of S consists of exactly k vectors.*

Theorem 4.8.2 provides a rigorous justification for using the term "dimension". For example, we have seen that $M_{2\times 2}$ has a basis that consists of exactly four vectors – the set of vectors $\{E_{11}, E_{12}, E_{21}, E_{22}\}$ from Exercise 4.8.2. This means that all bases of $M_{2\times 2}$ consist of exactly four vectors and thus we are justified in declaring that

$$\dim\big(M_{2\times 2}\big) = 4.$$

Similarly, we know that set $\{1, x, x^2, \ldots, x^n\}$ is a basis for $\mathbb{P}_n$. This basis consists of $n+1$ elements. Hence

$$\dim\big(\mathbb{P}_n\big) = n + 1.$$

4.9 Additional Exercises

(Jump to Solutions)

1. Prove statement a of Theorem 4.1.2. That is, show that any set of vectors in R^n that includes the zero vector, $\vec{0}_n$, is linearly dependent.

2. Prove that the set $\{\vec{0}_n\}$ is a subspace of R^n for any $n \geq 2$.

3. Prove that if S is a subspace of R^n, then S must contain the zero vector, $\vec{0}_n$.

4. Consider the subspace

$$S = \text{Span}\{\langle 1, 2, 1, 1\rangle, \langle 3, 5, 3, 4\rangle, \langle 1, 1, 1, 2\rangle, \langle 2, 1, 1, 4\rangle\}$$

of R^4. Find a basis for S.

5. Let $n \geq 2$; suppose S is a subspace of R^n and $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_k\}$ is a basis for S where $1 \leq k \leq n$. Explain why

$$[\vec{0}_n]_{\mathcal{B}} = \vec{0}_k.$$

That is, explain why the coordinate vector for the zero vector in S (which is the zero vector in R^n) must be the zero vector in R^k.

6. Determine whether the columns of A are linearly independent or linearly dependent. If the columns are linearly dependent, find a linear dependence relation.

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

   (b) $A = \begin{bmatrix} 1 & -2 & 0 & 1 \\ 2 & -4 & 2 & 6 \\ 0 & 0 & 3 & 6 \\ -3 & 6 & 1 & -1 \end{bmatrix}$

   (c) $A = \begin{bmatrix} 0 & 3 & 1 \\ 4 & 7 & 5 \\ -2 & -5 & -3 \\ 5 & -4 & 2 \end{bmatrix}$

   (d) $A = \begin{bmatrix} 3 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & -2 & 5 \\ -1 & 8 & 4 \end{bmatrix}$

7. Verify that $\mathbb{P}_n$ satisfies axioms 1–4 and axioms 7–10 of Definition 4.6.1. (Note that axioms 5 and 6 have already been discussed.)

8. For each statement, indicate whether the statement is true or false. Give a brief explanation of reason for each conclusion.

   (a) If A is an $n \times n$ matrix, then $\mathcal{RS}(A) = \mathcal{CS}(A)$.

   (b) If A is an $n \times n$ matrix, then $\dim\big(\mathcal{RS}(A)\big) = \dim\big(\mathcal{CS}(A)\big)$.

   (c) If A is a 3×3 matrix and $\text{rank}(A) = 3$, then the homogeneous equation $A\vec{x} = \vec{0}_3$ has only the trivial solution.

(d) The dimension of $\mathbb{P}_5$ is $\dim\big(\mathbb{P}_5\big) = 5$.

(e) If A is an $m \times n$ matrix with linearly dependent columns, then the equation $A\vec{x} = \vec{0}_m$ must have infinitely many solutions.

(f) If A is an $m \times n$ matrix with linearly dependent columns, then the equation $A\vec{x} = \vec{y}$ must have infinitely many solutions for any $\vec{y}$ in R^m.

(g) An element of a vector space is called a vector.

(h) If p is a vector in $\mathbb{P}_4$ and $\mathcal{B}$ is some basis of $\mathbb{P}_4$, then the coordinate vector $[p]_{\mathcal{B}}$ is a vector in R^5.

(i) For matrix A, if the first three rows of $\mathrm{rref}(A)$ are nonzero, then the first three rows of A are linearly independent.

(j) For matrix A, if the first three column vectors of $\mathrm{rref}(A)$ are three different standard unit vectors, then the first three columns of A are linearly independent.

9. For each matrix, find bases for the row space, the column space, and the null space.

(a) $A = \begin{bmatrix} 1 & 2 & -1 & 3 \\ 2 & 4 & -2 & 6 \\ 1 & 4 & 0 & 0 \end{bmatrix}$

(b) $B = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 5 & 0 \\ 4 & 6 & -2 \\ 2 & 3 & -1 \end{bmatrix}$

(c) $C = \begin{bmatrix} -2 & 2 & -3 & -2 & -8 \\ 3 & -3 & 3 & 1 & 10 \\ 2 & -2 & 2 & 0 & 4 \end{bmatrix}$

10. The first three *Chebyshev*[8] *polynomials* are

$$T_0(x) = 1, \quad T_1(x) = x, \quad \text{and} \quad T_2(x) = 2x^2 - 1.$$

[8]Pafnuty Lvovich Chebyshev (1821–1894) was a prominent Russian mathematician. Some transliterations of his name begins with the letter T, hence the convention of using T_i when naming the polynomials. The polynomials, more accurately called the Chebyshev polynomials of the first kind, arise as solutions to a specific family of differential equations and have applications in numerical analysis, signal processing, and other areas.

Show that the set $\mathcal{C} = \{T_0, T_1, T_2\}$ is a basis for $\mathbb{P}_2$.

11. Suppose A is a 7×10 matrix.

    (a) If $\mathcal{RS}(A)$ is a subspace of R^k, what is k?
    (b) If $\mathcal{CS}(A)$ is a subspace of R^k, what is k?
    (c) If $\mathcal{N}(A)$ is a subspace of R^k, what is k?
    (d) If $\text{rank}(A) = 7$, find $\text{nullity}(A)$.
    (e) If the homogeneous equation $A\vec{x} = \vec{0}_7$ has four free variables, what is $\dim(\mathcal{RS}(A))$?
    (f) If A is full rank, what is $\text{rank}(A)$? (Recall that full rank means that the rank is the largest it can be.)
    (g) If $\text{rank}(A) = 4$, find $\text{nullity}(A^T)$.
    (h) If $\text{rank}(A^T) = 6$, what is $\dim(\mathcal{CS}(A))$?

12. Which of the following sets, S, are subspaces of R^∞? Explain your answers.

    (a) $S = \text{Span}\left\{\langle 1, 3, 5, 7, \ldots\rangle, \langle 2, 4, 6, 8, \ldots\rangle\right\}$
    (b) $S = \{\vec{a} = \langle a_1, a_2, a_3, \ldots\rangle \in R^\infty \mid a_n \geq 0 \text{ for all } n = 1, 2, 3, \ldots\}$
    (c) $S = \{\vec{a} \in R^\infty \mid \vec{a} \text{ diverges}\}$ (*Note:* This question requires knowledge of Calculus II material.)
    (d) $S = \{\vec{a} \in R^\infty \mid \text{all entries of } \vec{a} \text{ are either 0 or 1 or } -1\}$
    (e) $S = \{\vec{a} \in R^\infty \mid \vec{a} \text{ has only finitely many non–zero entries}\}$

13. Let S be the set of all functions, f, in $C^1(R)$ that are equal to their derivative. In other words,

$$S = \left\{f \in C^1(R) \mid f' = f\right\}.$$

    (a) Which of the following functions, with domain R, are in the set S?

        i. $f(x) = x$

ii. $f(x) = x^2$
iii. $f(x) = e^x$
iv. $f(x) = 4e^x$
v. $f(x) = 7$
vi. $f(x) = \sin(x)$
vii. $f(x) = e^{3x}$

(b) Is S a subspace of $C^1(R)$? Explain.

14. Let T be the set of all functions, f, in $C^1(R)$ whose derivatives are equal to the function x^2. In other words,

$$S = \left\{ f \in C^1(R) \mid f'(x) = x^2 \text{ for all } x \in R \right\}.$$

(a) Which of the following functions, with domain R, are in the set S?

i. $f(x) = x$
ii. $f(x) = x^2$
iii. $f(x) = x^3$
iv. $f(x) = x^3 - 12$
v. $f(x) = \frac{1}{3}x^3 + 27$
vi. $f(x) = 3x^3$

(b) Is T a subspace of $C^1(R)$? Explain.

15. Let K be the set of all functions, f, in $C^0([-\pi, \pi])$ that satisfy

$$\int_{-\pi}^{\pi} f(x)\, dx = 0.$$

(a) Which of the following functions, with domain $[-\pi, \pi]$, are in the set K?

i. $f(x) = x$
ii. $f(x) = x^2$
iii. $f(x) = x^3$
iv. $f(x) = \sin(x)$
v. $f(x) = \cos(x)$

vi. $f(x) = x\sin(x)$

vii. $f(x) = x\cos(x)$

(b) Is K a subspace of $C^0([-\pi, \pi])$? Explain.

16. Let L be the set of all functions, f, in $C^0([-\pi, \pi])$ that satisfy

$$\int_{-\pi}^{\pi} f(x)\,dx = 1.$$

Is L a subspace of $C^0([-\pi, \pi])$? Explain.

17. Consider the set of functions

$$S = \{1, e^x\}$$

in $F(R)$.

Show that S is linearly independent and is thus a basis for Span(S).

Determine $[7 - 8e^x]_S$.

18. Consider the set

$$V = \{\langle x_1, x_2 \rangle \mid x_1 \in R \text{ and } x_2 \in R\}.$$

In other words, $V = R^2$, but we are going to define one of the operations on V differently than how we defined it for R^2 in Chapter 1. Thus V and R^2 are equal as sets, but not as vector spaces. In fact, the goal of this problem is to prove that V is not a vector space!

We will define addition of elements of V in the usual way: For any elements $\vec{x}$ and $\vec{y}$ in V we define

$$\vec{x} + \vec{y} = \langle x_1, x_2 \rangle + \langle y_1, y_2 \rangle = \langle x_1 + y_1, x_2 + y_2 \rangle.$$

However, we will define scalar multiplication in a different way: For any element $\vec{x} \in V$ and scalar $c \in R$ we define

$$c\vec{x} = \langle 0, 0 \rangle.$$

Explain why V with the operations defined as we have defined them above is not a vector space. This is a good exercise in understanding Definition 4.6.1. (Which of the ten axioms of Definition 4.6.1 does V satisfy and which of the axioms does it not satisfy?)

Chapter 5

Linear Transformations

The concept of *function* is a concept that you are familiar with from previous mathematics courses you have taken. In calculus, you studied a variety of functions such as polynomial functions, rational functions, trigonometric functions, exponential functions, and others. It is at this point in our linear algebra course that we introduce the function concept in the setting of vector spaces. However, in the vector space setting we will only consider special kinds of functions called linear transformations. Roughly speaking, these are functions that *map lines to lines or points.* A little less rough (but still rough) description of a linear transformation is as follows:

> A linear transformation, T, from a vector space V to a vector space W is a function $T : V \to W$ such that if L is any line in V, then applying T to L produces a line or a point in W.

Don't worry if you didn't understand the above statement when you first read it. It is just a vague definition that attempts to capture the overall idea of what a linear transformation is. You know what a vector space is, but the above statement will make sense to you only after we have given precise definitions of the terms "linear transformation" and "line", and you have had an opportunity to get familiar with these concepts as you proceed through this chapter and work on the exercises. After finishing your study of this chapter, you should come back and read the above statement again and hopefully it will make sense to you at that point!

5.1 General Ideas Pertaining to Functions

If V and W are vector spaces, then a linear transformation, T, from V to W, is a function that assigns each vector in V to some vector in W, but not every function from V to W is a linear transformation. A linear transformation is a function that satisfies two properties called *linearity properties* that are described in Definition 5.2.1 which is given in Section 5.2. The notation that we will use to say that T is a linear transformation from V to W is $T : V \to W$. This kind of notation is used when discussing functions in general (not just linear transformations) and we need to make sure that we understand it and some ideas and vocabulary related to it before we provide our definition of a linear transformation.

5.1.1 Domain, Codomain, Images, and Range

In general, if D and C are two non–empty sets and we write $f : D \to C$, then we are saying that f is a function whose *domain* is the set D and whose *codomain* is the set C. This means that f takes *inputs* from the set D and assigns them (by some rule) to *outputs* in the set C. So when you see the notation $f : D \to C$, here is how you should interpret it:

$$\underbrace{f}_{\text{Rule}} : \underbrace{D}_{\text{set of all inputs}} \longrightarrow \underbrace{C}_{\text{set where outputs live}} . \tag{5.1}$$

If x is any element of the domain D, then $f(x)$ denotes the element of C that x is assigned to by f. In other words, if x is some element of D and the function f assigns x to y, then we write $f(x) = y$. Instead of using the word "assign" we can use the word "map". Hence we can either say that f *assigns* x to y or we can say that f *maps* x to y. If $f(x) = y$, then we refer to y as the **image of x under f**. This is illustrated by the schematic diagram shown in Figure 5.1.

More generally, if S is any subset of the domain D, then we use the notation $f(S)$ to denote the set of all images of the elements of S and we call this the **image of S under f**. To write this formally (using set notation), the image of a set S under f is defined to be

$$f(S) = \{y \in C \mid y = f(x) \text{ for at least one } x \in S\}.$$

Written in a more compact form,

$$f(S) = \{f(x) \mid x \in S\}. \tag{5.2}$$

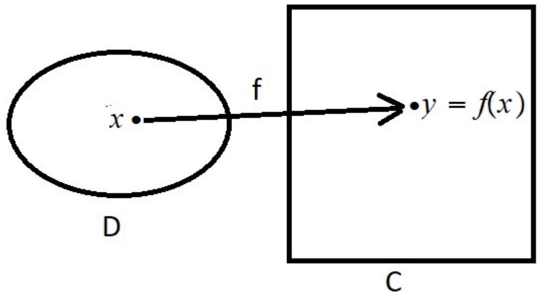


Figure 5.1: y is the image of x under f.

The schematic diagram shown in Figure 5.2 shows a subset, S, in the domain D and its image, $f(S)$, in the codomain C.

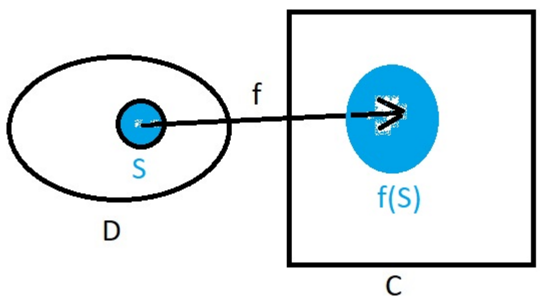


Figure 5.2: $f(S)$ is the image of S under f.

In writing the description (5.1), it was not by accident that we said that D is the "set of **all** inputs" and that C is the "set where outputs **live**". Depending on how we specify the codomain, C, it may or may not be the case that C is the set of all outputs of the function f. The subset of C that is the set of **all outputs** of the function f is what we call the **range** of f. The range of f is denoted by Range(f). A formal definition of Range(f) is

$$\text{Range}(f) = \{y \mid y = f(x) \text{ for at least one } x \in D\}.$$

An equivalent (and more compact) way to write the definition of Range(f) is

$$\text{Range}(f) = \{f(x) \mid x \in D\}.$$

In light of (5.2), an even more compact way to define Range(f) is

$$\text{Range}(f) = f(D).$$

So when thinking about the concept of range, here is what you should be thinking:

$$\underbrace{f}_{\text{Rule}} : \underbrace{D}_{\text{set of all inputs}} \longrightarrow \underbrace{\text{Range}(f)}_{\text{set of } \textbf{all outputs}} \tag{5.3}$$

The schematic diagram in Figure 5.3 illustrates Range$(f) = f(D)$.

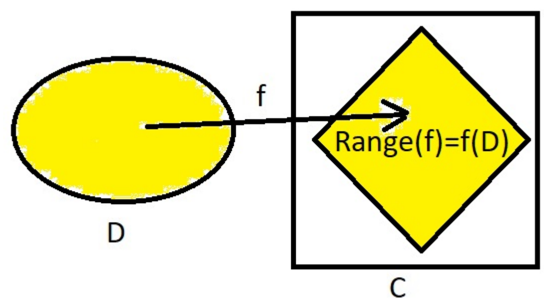


Figure 5.3: $f(D)$ is the image of D under f.

You may be asking yourself why it is even necessary to bother with the idea of a codomain. Once we have specified the domain D and the rule f, why don't we just simply write $f : D \to \text{Range}(f)$ as in (5.3)? Indeed it is correct to write this. However, when working with functions, it is often the case that the range of the particular function that is being investigated is not immediately obvious. Hence we choose some set, C, that we are sure is large enough to guarantee that all of the function outputs are contained in C, and we call this set the codomain. There is not a unique way to specify the codomain. We just need to be sure that it is a set that contains all of the outputs of the function. Furthermore, many investigations that arise in both theoretical and applied problems center around a whole class of functions (not just one function) that all have the same domain but have different ranges. But even though all of these functions have different ranges, the ranges are all contained in some common set C. An example of this type of situation was seen in Section 4.7.4.1 where we discussed the set $F(D)$ of *all* real–valued functions that have some common domain D. When we stipulated that we were considering *real–valued* functions, what we were actually doing was stipulating that the codomain of each of the functions being considered was understood to be the set of all real numbers R. The functions in $F(D)$ are all real–valued functions but they do not all have the same range. As a specific example, consider the functions $f : R \to R$ and $g : R \to R$ defined, respectively, by

$$f(x) = 3x$$
$$g(x) = x^2.$$

Both f and g have domain R and codomain R. However, the ranges of these functions are

$$\text{Range}(f) = f(R) = R$$
$$\text{Range}(g) = g(R) = [0, \infty).$$

Exercise 5.1.1. *Determine the range of each of the following functions $f : R \to R$. This exercise requires that you remember some things about elementary functions that you studied in calculus. All of these functions are continuous and thus the range of each function is an interval. You should try to do this exercise without using technology but it is OK if you just graph the given function using technology and determine the range by looking at the graph.*

The goal of this linear algebra course is not to study the types of functions that are contained in this exercise. (That is done in calculus.) It is OK if you skip this exercise (as long as your Instructor says it is OK). The exercise has been included here to help you to realize that the concept of Range is one that applies to functions in general - not just the functions we will study in linear algebra.

The first one is done as an example.

1. $f(x) = \sqrt{x^2 + 1}$

   Solution: *The function $y = x^2$ is an upward–opening parabola with vertex at the point $(0, 0)$. This means that x^2 ranges over the interval $[0, \infty)$ as x ranges over the interval $(-\infty, \infty)$.*

   The function $y = x^2 + 1$ is a vertical translation of the function $y = x^2$. Specifically, it is a translation in the upward direction by 1 unit. Thus $x^2 + 1$ ranges over the interval $[1, \infty)$ as x ranges over the interval $(-\infty, \infty)$.

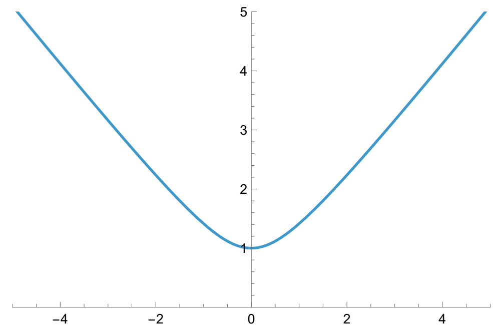


Figure 5.4: Graph of $f(x) = \sqrt{x^2+1}$

If we take the square root of any number in the interval $[1, \infty)$, *then we obtain another number in the interval* $[1, \infty)$. *In particular* $\sqrt{1} = 1$ *and if* x *is any number greater than* 1, *then* $\sqrt{x} > 1$. *Thus* $\sqrt{x^2+1}$ *ranges over the interval* $[1, \infty)$ *as* x *ranges over the interval* $(-\infty, \infty)$.

We conclude that $\text{Range}(f) = [1, \infty)$. *The graph of* f *in Figure 5.4.*

2. $f(x) = 4x$

3. $f(x) = 27$

4. $f(x) = x^2 - 6$

5. $f(x) = e^x$

6. $f(x) = \cos(x)$

7. $f(x) = |\cos(x)|$

8. $f(x) = 1 - x^2$

9. $f(x) = |1 - x^2|$

10. $f(x) = \sqrt{|1 - x^2|}$

5.1.2 Concepts of Onto, One-to-One, and Invertibility

If f is a function with domain D and designated codomain C, then we say that f maps D **into** C. If it happens to be the case that $\text{Range}(f) = C$

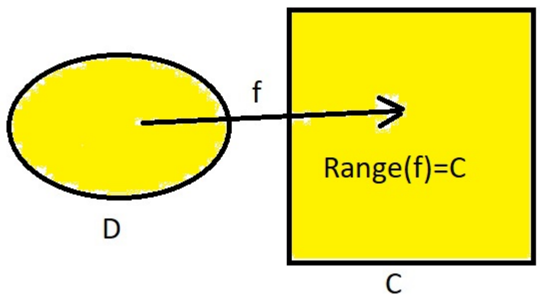


Figure 5.5: f maps D onto C.

(i.e., $f(D) = C$), then we say that f maps D **onto** C. The situation where f maps D onto C is illustrated in the schematic diagram in Figure 5.5.

As a specific example to illustrate the into/onto idea, once again consider the functions $f : R \to R$ and $g : R \to R$ defined by

$$f(x) = 3x$$
$$g(x) = x^2.$$

Both of these functions map R into R. f maps R onto R because $f(R) = R$. However, g does not map R onto R because $g(R) = [0, \infty) \neq R$.

In addition to the "onto" concept, another important concept pertaining to functions is the concept of a function being one–to–one. We have defined $\text{Range}(f) = f(D)$ and this means that $\text{Range}(f)$ is the set of all elements, y, in the codomain C such that there is at least one element $x \in D$ such that $f(x) = y$. Depending on what function, f, we are dealing with, it could be the case that there is some element $y \in \text{Range}(f)$ for which there is more than one $x \in D$ such that $f(x) = y$. If this is the case, then we say that the function f is **not one–to–one**. A specific example of a function that is not one–to–one is the function $g : R \to R$ defined by $g(x) = x^2$. To see why g is not one–to–one, just observe that $g(2) = 4$ and $g(-2) = 4$. The element 4 is in $\text{Range}(g) = [0, \infty)$ but there are two different elements (2 and -2) in the domain of g such that $g(x) = 4$. The schematic diagram in Figure 5.6 illustrates a function that is not one–to–one.

To summarize what we have said in the preceding paragraph: if $f : C \to D$ is a function and if there is at least one element $y \in \text{Range}(f)$ such that $f(x_1) = y$ and $f(x_2) = y$ for two *different* elements x_1 and x_2, in D, then we say that f is not one–to–one. We therefore say that f **is one–to–one**

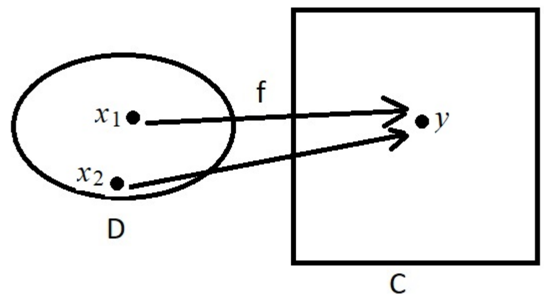


Figure 5.6: f is not one-to-one.

if for each element $y \in \text{Range}(f)$ there is a *unique* element x in D such that $f(x) = y$. As a specific example, the function $f : R \to R$ defined by $f(x) = 3x$ is one–to–one because for any element y in $\text{Range}(f) = R$, there is a unique element x in the domain R such that $3x = y$. For example, if we take the element $y = 8$ then the only element of R such that $f(x) = 8$ is $x = 8/3$ because the only solution of the equation $3x = 8$ is $x = 8/3$. Contrast this with the function $g : R \to R$ defined by $g(x) = x^2$, which is not one–to–one: If we take the element $y = 8 \in R$, then there is more than one solution in R of the equation $x^2 = 8$. This equation has two different solutions in R, which are $x_1 = \sqrt{8}$ and $x_2 = -\sqrt{8}$.

We can nicely summarize the foregoing discussion regarding the concepts of "onto" and "one–to–one" with the following definition.

Definition 5.1.1. *Suppose that D and C are non–empty sets and suppose that $f : D \to C$.*

1. *We say that the function f maps D* ***onto*** *C if for each element $y \in C$, the equation $f(x) = y$ has* ***at least one*** *solution in D.*

2. *We say that the function f is* ***one–to–one*** *if for each element $y \in \text{Range}(f)$ the equation $f(x) = y$ has a* ***unique*** *solution in D.*

Here is another way to interpret Definition 5.1.1 in words, using some other vocabulary words we have already encountered earlier in this course:

1. The function $f : D \to C$ maps D **onto** C if for each element $y \in C$, the equation $f(x) = y$ is consistent.

2. The function $f : D \to C$ is **one–to–one** if, assuming the equation $f(x) = y$ is consistent, it has a unique solution.

Functions that are both onto and one–to–one are said to be **invertible**. If $f : D \to C$ is both onto and one–to–one, then for **every** element $y \in C$ there is a **unique** element $x \in D$ such that $f(x) = y$. In other words, f is invertible if for all $y \in C$, the equation $f(x) = y$ is consistent **and** has a unique solution. If $f : D \to C$ is invertible, then there is a corresponding function, called the inverse of f and denoted by f^{-1} that "undoes what f does". Specifically $f^{-1} : C \to D$ is the function defined by

$$f^{-1}(y) = x \text{ where } x \text{ is the unique solution of } f(x) = y.$$

We can define invertibility using a definition that is in the spirit of Definition 5.1.1.

Definition 5.1.2. *Suppose that D and C are non–empty sets and suppose that $f : D \to C$. We say that f is* ***invertible*** *if for each element $y \in C$, the equation $f(x) = y$ has a unique solution.*

The inverse concept is illustrated in the schematic diagram in Figure 5.7.

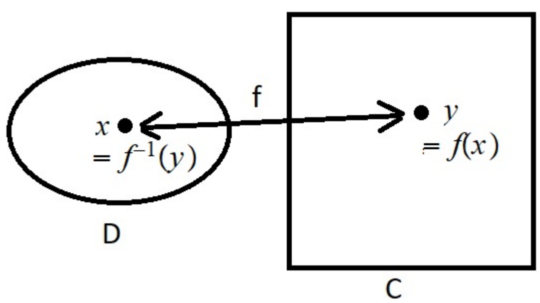


Figure 5.7: Illustration of f^{-1}.

The inputs of f^{-1} are the outputs of f and the outputs of f^{-1} are the inputs of f. When a certain x is put in to the function f, it produces the output $f(x)$. Then when this $f(x)$ is put in to the function f^{-1}, it produces the original x. This is summarized as follows: If $f : D \to C$ is invertible, then

$$\text{For all } x \in D,\ f^{-1}(f(x)) = x$$

and

$$\text{For all } x \in C,\ f\left(f^{-1}(x)\right) = x.$$

Also, observe that if f is invertible, then the domain of f^{-1} is the range of f and the range of f^{-1} is the domain of f.

5.1.3 Examples and Exercises

We will end this section with some examples illustrating the onto, one–to–one, and invertibility concepts and some exercises that should help you master these concepts. Our first two examples will use functions from calculus, which are non–linear functions, but then we will switch to examples that use linear functions from a vector space R^n into a vector space R^m. (We have not yet defined what the terms "non–linear function" and "linear function" mean, but we will do that in the upcoming section.) After these first two examples, and for the rest of this chapter, we will not give any more examples involving non–linear functions. You will probably find that the examples involving the non–linear functions are "harder" than the ones involving the linear functions. The reason is that the concepts of onto, one–to–one, and invertible involve studying equations of the form $f(x) = y$ and these equations can be difficult to study if f is a non–linear function such as $f(x) = \sin(x)$. However these equations are much easier to study if f is a linear function such as $f(x) = 3x$. In fact, one of our main focuses in this course so far has been to study equations of the type $f(\vec{x}) = \vec{y}$ where the function f is defined by $f(\vec{x}) = A\vec{x}$ where A is some given matrix. For those who wish to skip Examples 5.1.1 and 5.1.2 and Exercises 5.1.2 and 5.1.3 (assuming your instructor says it is OK to skip them!), you can resume reading with Example 5.1.3. We have chosen to include Examples 5.1.1 and 5.1.2 to emphasize that the concepts of onto, one–to–one, and invertibility apply much more broadly beyond linear algebra.

Example 5.1.1. *Consider the function $f : R \to R$ defined by $f(x) = x^2$. This function has domain R and has range*

$$\text{Range}(f) = f(R) = [0, \infty).$$

f does not map R onto R because Range$(f) \neq R$. *In addition, f is not one–to–one because $f(2) = f(-2) = 4$. (Thus two different inputs of f give the same output.)*

We will now "fix up" the definition of f by changing the domain and codomain, keeping the formula the same. This will result in the "new" function f being invertible.

Let f be the function $f : [0, \infty) \to [0, \infty)$ defined by $f(x) = x^2$. This function has domain $[0, \infty)$ and has range

$$\text{Range}(f) = f([0, \infty)) = [0, \infty).$$

In addition f is both onto $[0, \infty)$ and one–to–one. It is onto $[0, \infty)$ because for any choice of $y \in [0, \infty)$, the equation $x^2 = y$ has a solution in $[0, \infty)$. In fact such a solution is unique. For example, the unique solution of the equation $x^2 = 9$ that lies in the domain of f is $x = 3$. Since $x^2 = y$ has a unique solution for each y in Range(f), then f is invertible. The inverse of f is what we call the square root function. It is the function

$$f^{-1}(x) = \sqrt{x}.$$

$$\textit{Writing}\sqrt{3^2} = 3 \textit{ is the same as writing } f^{-1}(f(3)) = 3.$$

$$\textit{Writing } \left(\sqrt{3}\right)^2 = 3 \textit{ is the same as writing } f\left(f^{-1}(3)\right) = 3.$$

Example 5.1.2. *Consider the function $f : \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \to [-1, 1]$ defined by $f(x) = \sin(x)$. This function has domain $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ and has range*

$$\text{Range}(f) = f\left(\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\right) = [-1, 1].$$

We have purposely chosen the domain and codomain to make f be both onto its codomain and one–to–one. By looking at the graph of f which is shown in Figure 5.8, it can be seen that for every y in the codomain $[-1, 1]$, there is a unique x in the domain $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ such that $f(x) = y$. This tells us that f is invertible.

The inverse of this function is the inverse sine function which is denoted either by $\sin^{-1}$ or by arcsin. As a specific example, since

$$\sin\left(\frac{\pi}{3}\right) = \frac{\sqrt{3}}{2},$$

then

$$\sin^{-1}\left(\frac{\sqrt{3}}{2}\right) = \frac{\pi}{3}.$$

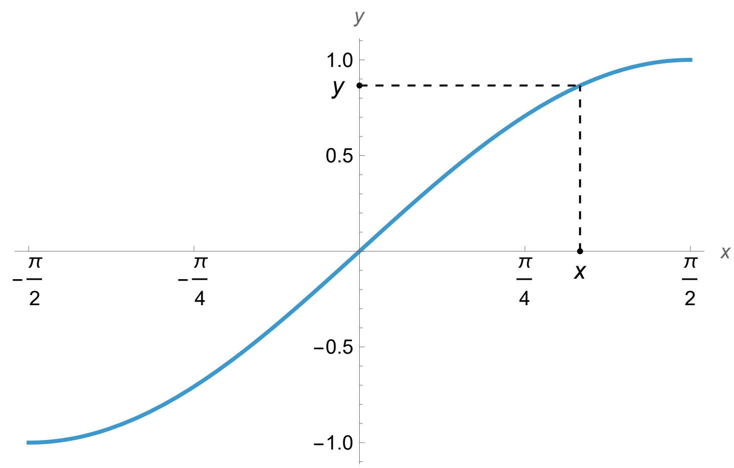


Figure 5.8: $f(x) = \sin(x)$.

Note that it is also true that

$$\sin\left(\frac{2\pi}{3}\right) = \frac{\sqrt{3}}{2},$$

but is is ***not true*** *that*

$$\sin^{-1}\left(\frac{\sqrt{3}}{2}\right) = \frac{2\pi}{3}.$$

That is because the number $2\pi/3$ *does not lie in the in the interval* $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$*. There is only one number,* x*, in the interval* $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ *for which* $\sin(x) = \sqrt{3}/2$ *and that number is* $x = \pi/3$*.*

Exercise 5.1.2. *In calculus (or perhaps in a precalculus course), you learned that the inverse of an exponential function is a logarithm function. This exercise should help you to understand the process of deriving that idea.*

Let $f : R \to (0, \infty)$ *be the function* $f(x) = e^x$*.*

1. *Draw the graph of* f *(either by hand or using technology).*

2. *Do you see that* f *maps* R *onto the interval* $(0, \infty)$ *and that* f *is one–to–one?*

3. *The inverse of* f *is called the natural logarithm function and is denoted by "*$\ln$*" or "*$\log_e$*". What is the domain of* f^{-1}*? What is the range of* f^{-1}*?*

4. *Fill in the blanks:*

$$\ln(e^x) = x \text{ for all } x \in \_\_\_\_\_\_\_\_$$

and

$$e^{\ln(x)} = x \text{ for all } x \in \_\_\_\_\_\_\_\_\_.$$

Exercise 5.1.3. *Let $f : D \to R$ be defined by the formula*

$$f(x) = \frac{1}{x}.$$

1. *What is the largest possible subset of R that you can choose for the domain, D, such that this formula makes sense (meaning that the formula produces a real number for all $x \in D$)?*
2. *Draw the graph of f (either by hand or using technology).*
3. *Using the domain, D, that you designated in part 1, what is* $\text{Range}(f)$*?*
4. *Do you see that f maps D onto its range and that f is one–to–one?*
5. *Find the formula for f^{-1}.*

Example 5.1.3. *Let $f : R \to R$ be the function defined by $f(x) = 5x$. If we take any number y in the codomain R, then the equation*

$$5x = y$$

has a unique solution in the domain R. Thus f is invertible.

The unique solution of $5x = y$ is

$$x = \frac{1}{5}y$$

and this tells us that the inverse of f is the function $f^{-1} : R \to R$ defined by

$$f^{-1}(x) = \frac{1}{5}x.$$

Here is a check that this is correct: For any $x \in R$ we have

$$f^{-1}(f(x)) = \frac{1}{5}f(x) = \frac{1}{5}(5x) = x$$

and

$$f\left(f^{-1}(x)\right) = 5f^{-1}(x) = 5\left(\frac{1}{5}x\right) = x.$$

Clearly, if we take any constant a *with* $a \neq 0$*, then the function* $f : R \to R$ *defined by* $f(x) = ax$ *is invertible and the inverse of* f *is the function* $f^{-1} : R \to R$ *defined by*

$$f^{-1}(x) = \frac{1}{a}x.$$

What if $a = 0$*? The function* $f : R \to R$ *defined by* $f(x) = 0x$ *does not map* R *onto* R *because* $\text{Range}(f) = \{0\} \neq R$*. In addition,* f *is not one–to–one because, for example* $f(12) = 0$ *and* $f(-37) = 0$*.*

Example 5.1.4. *The previous example considered functions* $f : R \to R$ *that have the form* $f(x) = ax$ *where* a *is a given constant. Throughout the rest of this chapter, we will consider functions of this form, except where* A *is a matrix! This is our first example.*

Let A *be the matrix*

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

and let $T : R^2 \to R^2$ *be the function defined by*

$$T(\vec{x}) = A\vec{x}.$$

Some questions we would like to answer are

1. *What is* $\text{Range}(T)$*?*
2. *Does* T *map* R^2 *onto* R^2*?*
3. *Is* T *one–to–one?*
4. *Is* T *invertible and, if so, what is* T^{-1}*?*

We are going to answer question 4 first (because the answer to question 4 will provide the answers to all of the other questions). We know that T *is invertible if any only if the equation*

$$A\vec{x} = \vec{y}$$

has a unique solution for every choice of $\vec{y}$ *in the codomain* R^2*. We know that this will be true if and only if both every row and every column of* A *has* *a pivot. Since A is a square matrix (size 2×2), this will be true if and only if* $\text{rref}(A) = I_2$*. It is easy to check that it is indeed true that* $\text{rref}(A) = I_2$*. Hence T is invertible.*

Since T is invertible, then T maps R^2 onto R^2 (which tells us that $\text{Range}(T) = R^2$*) and T is also one–to–one.*

Can you guess what the formula for T^{-1} is?

In Example 5.1.3, we saw that if $a \neq 0$, then the inverse of the function $f(x) = ax$ is the function $f^{-1}(x) = \frac{1}{a}x = a^{-1}x$. Perhaps not surprisingly, the inverse of $T(\vec{x}) = A\vec{x}$ is

$$T^{-1}(\vec{x}) = A^{-1}\vec{x}.$$

To see why this is so, note that for any $\vec{x}$ in R^2, we have

$$T^{-1}(T(\vec{x})) = A^{-1}(A\vec{x}) = \left(A^{-1}A\right)\vec{x} = I_2\vec{x} = \vec{x}$$

and likewise

$$T\left(T^{-1}(\vec{x})\right) = \vec{x}.$$

Example 5.1.5. *Let A be the matrix*

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

and let $T : R^2 \to R^2$ be the function defined by

$$T(\vec{x}) = A\vec{x}.$$

Some questions we would like to answer are

1. *What is* $\text{Range}(T)$*?*

2. *Does T map R^2 onto R^2?*

3. *Is T one–to–one?*

4. *Is T invertible and, if so, what is T^{-1}?*

The matrix A has reduced row echelon form and we see that

$$\text{rref}(A) = A \neq I_2.$$

Since not every row of A has a pivot, then there are some vectors $\vec{y}$ in R^2 for which $A\vec{x} = \vec{y}$ is inconsistent. This tells us that T does not map R^2 onto R^2.

Since not every column of A has a pivot, then even if $A\vec{x} = \vec{y}$ is consistent (which will be true if $\vec{y} \in \text{Range}\,(T)$), it has infinitely many solutions. This tells us that T is not one–to–one.

Clearly T is not invertible.

What is Range (T)? *We know that* Range (T) *is the set of all outputs of T. In other words,*

$$\text{Range}\,(T) = \left\{ A\vec{x} \mid \vec{x} \in R^2 \right\}.$$

Hence Range (T) *is precisely the column space of the matrix A. We know that the column space of A is the span of the pivot columns of A. Thus*

$$\text{Range}\,(T) = \text{Span}\,\{\langle\, 1, 0\rangle\}.$$

Example 5.1.6. *Let A be the matrix*

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

and let $T : R^3 \to R^2$ be the function defined by

$$T\,(\vec{x}) = A\vec{x}.$$

We can immediately see, due to the fact that A has more columns than rows, that it is not possible for the equation $A\vec{x} = \vec{y}$ to have a unique solution (even if $A\vec{x} = \vec{y}$ is consistent). Thus T is not one–to–one. This tells us that T is not invertible.

Since

$$\text{rref}\,(A) = \begin{bmatrix} 1 & 0 & -\frac{1}{9} \\ 0 & 1 & \frac{2}{3} \end{bmatrix},$$

we see that every row of A has a pivot. This tells us that T does map R^3 onto R^2.

Finally, since Range (T) *is the column space of A, and since we know that the column space of A is the span of the pivot columns of A, then*

$$\text{Range}\,(T) = \text{Span}\,\{\langle 3, 3\rangle, \langle 2, -1\rangle\}.$$

Example 5.1.7. *Let A be the matrix*

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

and let $T : R^2 \to R^3$ be the function defined by

$$T(\vec{x}) = A\vec{x}.$$

Since not every row of A has a pivot, then there are some $\vec{y} \in R^3$ for which $A\vec{x} = \vec{y}$ is inconsistent. This tells us that $\text{Range}(T) \neq R^3$ *and hence T does not map R^2 onto R^3. Hence T is not invertible. Also*

$$\text{rref}(A) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$

showing that all columns of A have a pivot. Therefore T is one–to–one.

The range of T is

$$\text{Range}(T) = \text{Span}\{\langle 2, -1, 0\rangle, \langle 0, 1, -2\rangle\}.$$

Exercise 5.1.4. *For each of the $m \times n$ matrices, A, given below, let $T : R^n \to R^m$ be the function defined by $T(\vec{x}) = A\vec{x}$. Answer each of the questions for each one. Explain your answers. Studying the above examples will help you see how you should explain your answers.*

a) What is the domain of T?

b) What is the codomain of T?

c) Does T map R^n onto R^m?

d) Is T one–to–one?

e) Is T invertible? If it is, then what is T^{-1}?

f) Describe the range of T as the span of some set of vectors in R^m.

1.

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

2.

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

3.

$$A = \begin{bmatrix} 0 & 3 \\ 0 & -4 \end{bmatrix}$$

4.

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

5.

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

6.

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

7.

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

8.

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

9.

$$A = \begin{bmatrix} 4 & 2 \end{bmatrix}$$

10.

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

Exercise 5.1.5. *Explain why if A is any $m \times n$ matrix and $T : R^n \to R^m$ is the function defined by $T(\vec{x}) = A\vec{x}$, then $\vec{0}_m \in \mathrm{Range}(T)$.*

Exercise 5.1.6. *Suppose that A is an $m \times n$ matrix and suppose that $T : R^n \to R^m$ is the function defined by $T(\vec{x}) = A\vec{x}$.*

Explain why if $m \neq n$, then T is not invertible.

Exercise 5.1.7. *Let $M_{2\times 2}$ be the set of all 2×2 matrices with real number entries. Let $f : M_{2\times 2} \to M_{2\times 2}$ be the function defined by $f(A) = \mathrm{rref}(A)$*

1. *Does f map $M_{2\times 2}$ onto $M_{2\times 2}$? Explain.*

2. *Is f one-to-one? Explain.*

5.2 Linear Transformations from R^n to R^m

For the rest of this chapter we will focus on a certain class of functions called linear transformations. These are functions whose domains and codomains are vector spaces and which satisfy two requirements which are given in Definition 5.2.1 below. We will start by focusing only on linear transformations that have domain R^n (for some n) and codomain R^m (for some m). In Section 5.6, we will generalize and study linear transformations whose domain and codomain can be any vector space. For notation, we will use capital letters such as T to denote linear transformations.

Definition 5.2.1. *A linear transformation from R^n to R^m is a function $T : R^n \to R^m$ that has the properties:*

1. *If $\vec{x}$ and $\vec{y}$ are any two vectors in R^n, then $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$.*

2. *If $\vec{x}$ is any vector in R^n and c is any scalar, then $T(c\vec{x}) = cT(\vec{x})$.*

Properties 1 and 2 of Definition 5.2.1 are usually referred to as **linearity properties**. Any function, T, that satisfies Properties 1 and 2 is called a **linear function** (we actually use the name "transformation" instead of "function" in linear algebra) and any function that does not satisfy Properties 1 and 2 is called a **nonlinear function**. (See Exercise 3 in the Additional Exercises, Section 5.8, for a nuance regarding use of the terms "linear" and "nonlinear".)

In the introductory section of this chapter (page 247), we stated a rather crude definition of what a linear transformation is – saying that a linear transformation is a function that maps lines to lines or points. Upon your first reading of Definition 5.2.1, it is probably not at all clear to you that this definition describes a kind of function that maps lines to lines or points. If you find the definition to be mysterious or as having come from "out of the blue", then that is normal. Any time we are introduced to a new mathematical concept and this introduction comes from reading a definition, we usually can't really grasp the definition until after we have had the opportunity to work with it using concrete examples. We will do that in Section 5.3. After reading some examples and working some exercises, you will hopefully be convinced that Definition 5.2.1 contains precisely the right ingredients (no more or no less) to define a class of functions that map lines to lines or points.

We have already seen some examples of linear transformations in Section 5.1.3. The functions that were studied in Examples 5.1.4, 5.1.5, 5.1.6, and 5.1.7, and the functions that were studied in Exercise 5.1.4 were all functions defined by formulas of the form $T(\vec{x}) = A\vec{x}$ where A was a given matrix. These are all linear transformations. The fact that any function of the form $T(\vec{x}) = A\vec{x}$ is a linear transformation is stated in the following lemma. The proof of the lemma relies on the matrix algebra properties that were studied in Chapter 3.

Lemma 5.2.1. *Suppose that A is an $m \times n$ matrix. Then the function $T : R^n \to R^m$ defined by $T(\vec{x}) = A\vec{x}$ is a linear transformation.*

Proof. In order to prove that T is a linear transformation, we need to show that $T(\vec{x}) = A\vec{x}$ satisfies both of the linearity properties stated in Definition 5.2.1.

First we will verify that Property 1 is satisfied: Let $\vec{x}$ and $\vec{y}$ be any two vectors in R^n. Then

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

Next we will verify that Property 2 is satisfied: Let $\vec{x}$ be a vector in R^n and let c be a scalar. Then

$$T(c\vec{x}) = A(c\vec{x}) = c(A\vec{x}) = cT(\vec{x}).$$

□

Lemma 5.2.1 tells us that every function of the form $T(\vec{x}) = A\vec{x}$, where A is a matrix, is a linear transformation. Our first big theorem regarding linear linear transformations is that the *only* linear transformations $T : R^n \to R^m$ are functions of the form $T(\vec{x}) = A\vec{x}$. The theorem also tells us how to find the matrix A.

Theorem 5.2.1. *Suppose that $T : R^n \to R^m$ is a linear transformation. Then there is a unique $m \times n$ matrix A, such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in R^n$.*

Furthermore, the matrix A is the matrix whose column vectors are

$$\mathrm{Col}_j(A) = T(\vec{e}_j)$$

where $\mathcal{E} = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ is the standard basis for R^n.

Proof. Suppose that $T : R^n \to R^m$ is a linear transformation.

Let $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle$ be some arbitrarily chosen vector in R^n.

Then

$$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n.$$

Since T is a linear transformation, we can use the linearity properties to compute $T(\vec{x})$. We obtain

$$\begin{aligned} T(\vec{x}) &= T(x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n) \qquad (5.4) \\ &= T(x_1\vec{e}_1) + T(x_2\vec{e}_2) + \cdots + T(x_n\vec{e}_n) \\ &= x_1T(\vec{e}_1) + x_2T(\vec{e}_2) + \cdots + x_nT(\vec{e}_n). \end{aligned}$$

Now define A to be the $m \times n$ matrix whose column vectors are

$$\mathrm{Col}_j(A) = T(\vec{e}_j). \qquad (5.5)$$

By looking at the calculation done in (5.4), we see that $T(\vec{x})$ is a linear combination of the column vectors of A using the entries of $\vec{x}$ as weights. Therefore $T(\vec{x}) = A\vec{x}$.

We have shown that the matrix A defined by (5.5) is such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in R^n$. We still need to prove that this matrix A is the only matrix for which this is true. Thus, suppose that B is some $m \times n$ matrix such that $T(\vec{x}) = B\vec{x}$ for all $\vec{x} \in R^n$. Then for any standard basis vector $\vec{e}_j$, $j = 1, 2, \ldots, n$, we have $T(\vec{e}_j) = B\vec{e}_j$ and thus

$$\mathrm{Col}_j(B) = B\vec{e}_j = T(\vec{e}_j) = \mathrm{Col}_j(A).$$

Since $\mathrm{Col}_j(B) = \mathrm{Col}_j(A)$ for all $j = 1, 2, \ldots, n$, then $B = A$, which is what we wanted to prove.

□

Theorem 5.2.1 really says a lot. It not only tells us that every linear transformation $T : R^n \to R^m$ has the form but $T(\vec{x}) = A\vec{x}$, but it also tells us how to find the matrix A. To find the matrix A, we only need to evaluate the linear transformation T at each of the basis vectors in the standard basis $\mathcal{E} = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$. This means that a linear transformation is *completely determined* by what it does to the standard basis vectors! We will refer to the unique matrix A such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in R^n$ as the **standard matrix of** T or simply as the **matrix of** T. Since A is uniquely determined by T, you may be wondering why we would ever bother to use the word "standard" when referring to this matrix. Why not just call the matrix A "the matrix" for T? There are no other matrices, B, such that $T(\vec{x}) = B\vec{x}$ for all $\vec{x} \in R^n$. The only matrix that works for this purpose is A. We can and will refer to A as "the matrix" of T, but the reason we may sometimes want to call A "the standard matrix" for T is because we use the standard basis, $\mathcal{E}$, of R^n to find A. As we will see, we can also use any basis, $\mathcal{B}$, of R^n (not just the standard basis) to determine a linear transformation, and in doing so we get another matrix that is related to T in a way that we will later make precise.

Example 5.2.1. *Let $T : R^2 \to R^2$ be the function defined by*

$$T(\langle x_1, x_2 \rangle) = \langle -x_2, x_1 \rangle . \tag{5.6}$$

Use Theorem 5.2.1 to find the matrix A such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in R^2$.

Solution: *Theorem 5.2.1 tells us that the matrix A is the matrix whose columns are $\text{Col}_j(A) = T(\vec{e}_j)$ where $\mathcal{E} = \{\vec{e}_1, \vec{e}_2\} = \{\langle 1, 0 \rangle, \langle 0, 1 \rangle\}$ is the standard basis for R^2. Using the formula (5.6), we obtain*

$$\begin{aligned} T(\vec{e}_1) &= T(\langle 1, 0 \rangle) = \langle -0, 1 \rangle = \langle 0, 1 \rangle \\ T(\vec{e}_2) &= T(\langle 0, 1 \rangle) = \langle -1, 0 \rangle . \end{aligned}$$

Thus the matrix for T is

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

To check that this is correct, we need to check that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in R^2$.

If we take any $\vec{x} = \langle x_1, x_2 \rangle$ in R^2, then

$$T(\vec{x}) = T(\langle x_1, x_2 \rangle) = \langle -x_2, x_1 \rangle$$

and

$$\begin{aligned} A\vec{x} &= x_1 \operatorname{Col}_1(A) + x_2 \operatorname{Col}_2(A) \\ &= x_1 \langle 0,1\rangle + x_2 \langle -1,0\rangle \\ &= \langle 0,x_1\rangle + \langle -x_2,0\rangle \\ &= \langle -x_2,x_1\rangle . \end{aligned}$$

This verifies that $T(\vec{x}) = A\vec{x}$ for all $\vec{x}$ in R^2.

Exercise 5.2.1. *Let $T : R^2 \to R^2$ be the function defined by*

$$T(\langle x_1,x_2\rangle) = \langle x_1 - x_2, 2x_1\rangle .$$

Use Theorem 5.2.1 to find the matrix A such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in R^2$.

Recall that if $A = [a_{ij}]$ is an $m \times n$ matrix and $\vec{x} = \langle x_1, x_2, \ldots, x_n\rangle$ is a vector in R^n, then

$$A\vec{x} = \langle \operatorname{Row}_1(A)\cdot\vec{x}, \operatorname{Row}_2(A)\cdot\vec{x}, \ldots . \operatorname{Row}_m(A)\cdot\vec{x}\rangle .$$

This tells us that if $T : R^n \to R^m$ is the linear transformation that has matrix A, then

$$T(\vec{x}) = \langle \operatorname{Row}_1(A)\cdot\vec{x}, \operatorname{Row}_2(A)\cdot\vec{x}, \ldots . \operatorname{Row}_m(A)\cdot\vec{x}\rangle .$$

Since, for any $i = 1, 2, \ldots, m$, we have

$$\operatorname{Row}_i(A)\cdot\vec{x} = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n,$$

we see that each entry of the vector $T(\vec{x})$ is a linear expression in the variables $x_1, x_2, \ldots, x_n$. This makes it easy to recognize which formulas give linear transformations and which do not. For example, just by glancing at the formula

$$T(\langle x_1,x_2\rangle) = \langle 2x_1 + 4x_2, -3x_1 + 7x_2\rangle ,$$

we can see that T is a linear transformation, because the components of $T(\vec{x})$ are defined to be $2x_1+4x_2$ and $-3x_1+7x_2$, which are both linear expressions in the variables x_1 and x_2. Likewise, we can just glance at the formula

$$T(\langle x_1,x_2\rangle) = \left\langle 2x_1 - 6x_2, -3x_1^2 + 5x_2\right\rangle ,$$

and see that this is not a linear transformation because the second component of $T(\vec{x})$ is $-3x_1^2 + 5x_2$ and this is not a linear expression in x_1 and x_2 (due to the fact that we see x_1^2 in the expression).

Exercise 5.2.2. *Determine whether or not each of the following expressions defines a linear transformation $T : R^n \to R^m$ (for appropriate m and n).*

1. $T(\langle x_1, x_2 \rangle) = \langle -4x_1 - 4x_2, -5x_1 + 3x_2 \rangle$
2. $T(\langle x_1, x_2, x_3 \rangle) = \langle x_1, -2x_2 + x_{3,}0 \rangle$
3. $T(\langle x_1, x_2 \rangle) = \langle x_1, 6x_2, -3x_1 \rangle$
4. $T(\langle x_1, x_2 \rangle) = \langle x_1, 6 \rangle$
5. $T(\langle x_1, x_2, x_3 \rangle) = \left\langle \sqrt{x_1^2 + x_2^2 + x_3^2}, 0 \right\rangle$
6. $T(\langle x_1, x_2 \rangle) = \langle 5x_1 - 3x_2^5, 9x_2 \rangle$
7. $T(\langle x_1, x_2 \rangle) = \langle 0, 0 \rangle$
8. $T(\langle x_1, x_2, x_3, x_4 \rangle) = \langle x_1 - x_2 + x_3 - 4x_4, x_1 - 2x_4, 3x_2 + x_3 \rangle$

Exercise 5.2.3. *Each of the functions $T : R^n \to R^m$ given below is a linear transformation (for some approriate values of n and m). Use Theorem 5.2.1 to find the matrix A such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in R^n$.*

1. $T(\langle x_1, x_2 \rangle) = \langle x_1, 0 \rangle$
2. $T(\langle x_1, x_2 \rangle) = \langle -x_1, x_2 \rangle$
3. $T(\langle x_1, x_2 \rangle) = \langle 2x_1, 2x_2 \rangle$
4. $T(\langle x_1, x_2 \rangle) = \langle 2x_1, -3x_2 \rangle$
5. $T(\langle x_1, x_2 \rangle) = \langle x_2, x_2 \rangle$
6. $T(\langle x_1, x_2 \rangle) = \left\langle \frac{1}{2}x_1 + \frac{1}{2}x_2, \frac{1}{2}x_1 + \frac{1}{2}x_2 \right\rangle$
7. $T(\langle x_1, x_2 \rangle) = \langle -4x_1, -2x_1 + 5x_2 \rangle$
8. $T(\langle x_1, x_2, x_3 \rangle) = \langle 8x_1 - 4x_2 + x_3, -4x_1 + 3x_3, 3x_1 + x_2 - 5x_3 \rangle$
9. $T(\langle x_1, x_2, x_3 \rangle) = \langle -7x_1 + 5x_3, -4x_1 + x_2 + 7x_3 \rangle$
10. $T(\langle x_1, x_2, x_3 \rangle) = \langle -8x_2 + x_3, -5x_1 + 2x_2 - 6x_3, 3x_1 + 6x_2 - 8x_3, -2x_1 + 7x_2 - 7x_3 \rangle$

We conclude this section with a theorem that states a simple basic property of all linear transformations $T : R^n \to R^m$.

Theorem 5.2.2. *If $T : R^n \to R^m$ is a linear transformation, then*

$$T\left(\vec{0}_n\right) = \vec{0}_m.$$

Proof. If $T : R^n \to R^m$ is a linear transformation, then we know by Theorem 5.2.1 that there is a (unique) $m \times n$ matrix A such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x}$ in R^n. Plugging in $\vec{x} = \vec{0}_n$, we obtain

$$T\left(\vec{0}_n\right) = A\vec{0}_n = \vec{0}_m.$$

□

The fact that $T\left(\vec{0}_n\right) = \vec{0}_m$ must be true for any linear transformation is important. For one thing, it tells us that the vector $\vec{0}_m$ is always in Range (T). In other words, the homogeneous equation $T(\vec{x}) = \vec{0}_m$ is always consistent because it has at least the trivial solution $\vec{x} = \vec{0}_n$. As you might suspect based on your experience so far in this course, it is a question of interest to ask whether or not the equation $T(\vec{x}) = \vec{0}_m$ has any non–trivial solutions. If it does, then the linear transformation T is not one–to–one and hence not invertible. We will consider this and related issues in the upcoming Section 5.2.1.

5.2.1 Fundamental Subspaces of Linear Transformations

In Section 4.2.1, we discussed the four fundamental subspaces of an $m \times n$ matrix, A. Two of these subspaces are the column space of A and the null space of A.

The column space of A, denoted by $\mathcal{CS}(A)$, is defined to be

$$\mathcal{CS}(A) = \text{Span}\left\{\text{Col}_1(A), \text{Col}_2(A), \ldots, \text{Col}_n(A)\right\}.$$

Since the column space of A is the span of the column vectors of A, then an equivalent way to define the column space of A is

$$\mathcal{CS}(A) = \left\{A\vec{x} \mid \vec{x} \in R^n\right\}. \tag{5.7}$$

The null space of A, denoted by $\mathcal{N}(A)$, is defined to be the set of all $\vec{x} \in R^n$ such that $A\vec{x} = \vec{0}_m$. Thus

$$\mathcal{N}(A) = \left\{\vec{x} \in R^n \mid A\vec{x} = \vec{0}_m\right\}. \tag{5.8}$$

$\mathcal{CS}(A)$ is a subspace of R^m and $\mathcal{N}(A)$ is a subspace of R^n and in Section 4.5, we presented the Fundamental Theorem of Linear Algebra, part of which tells us that the sum of the dimensions of $\mathcal{CS}(A)$ and $\mathcal{N}(A)$ must be n. Thus it is true for any $m \times n$ matrix, A, that

$$\dim(\mathcal{CS}(A)) + \dim(\mathcal{N}(A)) = n. \tag{5.9}$$

When we are considering a linear transformation $T : R^n \to R^m$, the concept that is analogous to column space is the concept of range! Recall that $\text{Range}(T)$ is defined to be

$$\text{Range}(T) = \{T(\vec{x}) \mid \vec{x} \in R^n\}. \tag{5.10}$$

By looking at the definition of column space given in (5.7), we see that if A is the matrix of T, meaning that $T(\vec{x}) = A\vec{x}$ for all $\vec{x}$ in R^n, then the definition of $\text{Range}(T)$ given in (5.10) is identical to (5.7). Thus

$$\text{Range}(T) = \mathcal{CS}(A).$$

The linear transformation concept that is analogous to the concept of the null space of a matrix is one that we have not yet defined, but it is what you may suspect. It is the set of all vectors $\vec{x} \in R^n$ such that $T(\vec{x}) = \vec{0}_m$. We call this set the kernel of T or the null space of T.

Definition 5.2.2. *For a linear transformation $T : R^n \to R^m$, the **kernel** of T (also called the **null space** of T) is defined to be*

$$\ker(T) = \left\{\vec{x} \in R^n \mid T(\vec{x}) = \vec{0}_m\right\}.$$

As can be seen by looking at (5.8), if A is the matrix such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in R^n$, then

$$\ker(T) = \mathcal{N}(A).$$

We know that $\mathcal{N}(A)$ is subspace of R^n and hence $\ker(T)$ (which is the same thing as $\mathcal{N}(A)$) is a subspace of R^n, and we can translate the second statement of the Fundamental Theorem of Linear Algebra (Theorem 4.5.1), which is the statement given in (5.9), to make a corresponding statement that applies to linear transformations.

Theorem 5.2.3. *Suppose that $T : R^n \to R^m$ is a linear transformation. Then*

$$\dim(\text{Range}(T)) + \dim(\ker(T)) = n.$$

In Section 5.4, we will prove that any linear transformation, T, maps lines to lines or points. We will see that the subspace $\ker(T)$ plays an interesting role in determining whether a particular line is mapped by T to a line or to a point. An important connection between $\text{Range}(T)$ and $\ker(T)$ and the concepts of onto and one–to–one is given in the following corollary.

Corollary 5.2.1. *Suppose that $T : R^n \to R^m$ is a linear transformation. Then*

1. *T maps R^n onto R^m if and only if $\dim(\text{Range}(T)) = m$ and*
2. *T is one to one if and only if $\dim(\ker(T)) = 0$.*

Proof. Suppose that T maps R^n onto R^m. This means that $\text{Range}(T) = R^m$ and we conclude that

$$\dim(\text{Range}(T)) = \dim(R^m) = m.$$

Conversely, suppose that $\dim(\text{Range}(T)) = m$. Since $\text{Range}(T)$ is a subspace of R^m and since the only subspace of R^m that has dimension m is R^m itself, then $\text{Range}(T) = R^m$, which means that T maps R^n onto R^m.

Suppose that T is one to one. Then the equation $T(\vec{x}) = \vec{0}_n$ has only the trivial solution $\vec{x} = \vec{0}_n$. This means that $\ker(T) = \left\{\vec{0}_n\right\}$ and hence $\dim(\ker(T)) = 0$.

Conversely, suppose that $\dim(\ker(T)) = 0$. Then, by Theorem 5.2.3, we know that $\dim(\text{Range}(T)) = n$. Since the matrix, A, that defines T is an $m \times n$ matrix and $\text{Range}(T)$ is the column space of A, then the column space of A has dimension n, which means that every column of A is a pivot column and hence T is one–to–one. □

We now provide an example and some exercises that illustrate Theorem 5.2.3.

Example 5.2.2. *Let A be the linear transformation defined by*

$$T(\langle x_1, x_2 \rangle) = \langle x_1 + 3x_2, 3x_1 + 9x_2 \rangle.$$

We will illustrate Theorem 5.2.3 for this transformation.

First we note that the

$$T(\langle 1,0\rangle) = \langle 1,3\rangle$$
$$T(\langle 0,1\rangle) = \langle 3,9\rangle$$

and thus the matrix of T is

$$A = \begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix}.$$

Since

$$\operatorname{rref}(A) = \begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix},$$

we see that

$$\operatorname{Range}(T) = \mathcal{CS}(A) = \operatorname{Span}\{\langle 1,3\rangle\}.$$

In addition we see that

$$\ker(T) = \mathcal{N}(A) = \operatorname{Span}\{\langle -3,1\rangle\}.$$

Hence

$$\dim(\operatorname{Range}(T)) + \dim(\ker(T)) = 1 + 1 = 2.$$

Exercise 5.2.4. *Illustrate Theorem 5.2.3 for the linear transformations $T : R^n \to R^m$ given in 1-5. Specifically,*

a) *Find the matrix, A, of T.*

b) *Find* $\operatorname{Range}(T)$ *and write it in the form* $\operatorname{Range}(T) = \operatorname{Span}\{\textit{basis vectors}\}$.

c) *Find* $\ker(T)$ *and write it in the form* $\ker(T) = \operatorname{Span}\{\textit{basis vectors}\}$.

d) *Verify that* $\dim(\operatorname{Range}(T)) + \dim(\ker(T)) = n$.

1. $T(\langle x_1, x_2\rangle) = \langle 3x_1 + 4x_2, 4x_2\rangle$
2. $T(\langle x_1, x_2\rangle) = \langle x_1, 5x_1\rangle$
3. $T(\langle x_1, x_2, x_3\rangle) = \langle x_1 + x_2 + x_3, x_1 + x_2 + x_3, x_1 + x_2 + x_3\rangle$
4. $T(\langle x_1, x_2, x_3, x_4\rangle) = \langle 5x_1 + 7x_2 - 3x_3 - 2x_4, 6x_1 + 2x_2 + x_3 + 2x_4\rangle$

5. $T(\langle x_1, x_2, x_3\rangle) = \langle 0,0,0,0,0\rangle$.

Exercise 5.2.5. *Suppose that* $T : R^6 \to R^3$ *is a linear transformation. Explain why it is not possible that* $\ker(T) = \left\{\vec{0}_6\right\}$.

Exercise 5.2.6. *Suppose that* $T : R^4 \to R^5$ *is a linear transformation. Explain why it is not possible that* $\text{Range}(T) = R^5$.

Exercise 5.2.7. *Suppose that* $E : R^4 \to R^4$ *is the* ***identity transformation****, which is defined by* $E(\vec{x}) = \vec{x}$ *for all* $\vec{x} \in R^4$*. What are the dimensions of* $\text{Range}(E)$ *and* $\ker(E)$*?*

Exercise 5.2.8. *Suppose that* $Z : R^4 \to R^4$ *is the* ***zero transformation****, which is defined by* $Z(\vec{x}) = \vec{0}_4$ *for all* $\vec{x} \in R^4$*. What are the dimensions of* $\text{Range}(Z)$ *and* $\ker(Z)$*?*

5.3 Visualizing Linear Transformations

Our understanding of a function is enhanced if we can draw a picture of the function. For example. if we just say that f is the function

$$f(x) = \sin(x),$$

then all we see is a formula. But if we draw the graph of this function (shown in Figure 5.9), then the function comes alive.

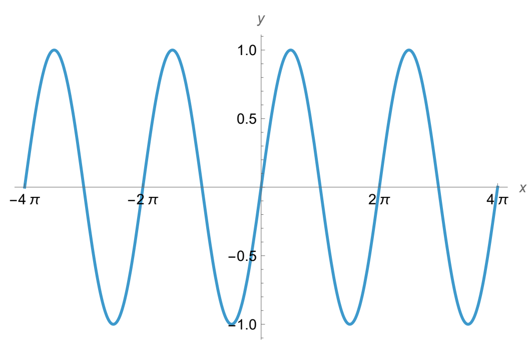


Figure 5.9: $f(x) = \sin(x)$

The graph gives us the “big picture” of $f(x) = \sin(x)$. In a glance, we can see that the maximum value of the function is 1 and that the minimum value if -1. We can see that the graph of the function is smooth (differentiable) and that it oscillates repeatedly between the values 1 and -1. We can see that the graph crosses the x axis infinitely many times and hence we know that the equation $\sin(x) = 0$ has infinitely many solutions in R. As the old saying goes, "a picture is worth a thousand words."

Linear transformations are functions, and at the beginning of this chapter we introduced the topic of linear transformations by saying that linear transformations are functions that map lines to lines or points. "Lines" and "points" are things we can draw pictures of, so might we somehow be able to draw a picture of a linear transformation? If so, then that should help us better understand how linear transformations behave, just as drawing the graph of $f(x) = \sin(x)$ helps us to understand how that function behaves. The good news is that we *can* draw pictures that help us understand the behavior of linear transformations but, as experience has taught us, we should probably not try to stretch out artistic abilities too far. We should probably limit our efforts to trying to draw pictures of linear transformations $T : R^2 \to R^2$. That is what we will do in this section. So let us now embark on the effort to make some linear transformations "come alive".

Something that we are going to find to be useful in drawing pictures of linear transformations $T : R^2 \to R^2$ is Theorem 5.2.1, which tells us that such a linear transformation is completely determined by its action on the standard basis vectors $\vec{e}_1 = \langle 1, 0\rangle$ and $\vec{e}_2 = \langle 0, 1\rangle$. If we are given $T(\vec{e}_1)$ and $T(\vec{e}_2)$, then we immediately know that the formula for T is

$$T(\vec{x}) = A\vec{x} \text{ for all } \vec{x} = \langle x_1, x_2\rangle \text{ in } R^2$$

where A is the 2×2 matrix such that

$$\begin{aligned} \mathrm{Col}_1(A) &= T(\vec{e}_1) \\ \mathrm{Col}_2(A) &= T(\vec{e}_2). \end{aligned}$$

In the examples we are about to present, we will make use of this fact, but we will also employ the more elementary "hands–on" approach of just choosing some selected vectors and plotting their standard representatives along with the standard representatives of their images under T. Studying these examples and working some exercises should help you get more comfortable with linear transformations. In particular, you should begin to see why the word "transformation" is a good word to describe what these functions do. They "transform" vectors in R^2 into other vectors in R^2.

5.3.1 Example: A Rotation Transformation

We will begin by studying the linear transformation $T : R^2 \to R^2$ defined by

$$T(\langle x_1, x_2 \rangle) = \langle -x_2, x_1 \rangle . \tag{5.11}$$

Let us get comfortable with the basic idea of how we can start drawing a picture of this function by picking vectors in the domain and plotting them along with their images in the range.

If we (randomly) choose the vectors $\langle 1, 0 \rangle$ and $\langle 4, -2 \rangle$ in the domain, we see that their images under T are

$$T(\langle 1, 0 \rangle) = \langle 0, 1 \rangle$$

and

$$T(\langle 4, -2 \rangle) = \langle 2, 4 \rangle$$

We can draw a picture illustrating the fact that $T(\langle 1, 0 \rangle) = \langle 0, 1 \rangle$ by drawing the standard representative of the input $\langle 1, 0 \rangle$ in black and drawing the standard representative of the output $\langle 0, 1 \rangle$ in red. (Or you may choose other colors if these are not your favorites.) Likewise, we can illustrate the fact that $T(\langle 4, -2 \rangle) = \langle 2, 4 \rangle$ by drawing the standard representative of the input $\langle 4, -2 \rangle$ in black and drawing the standard representative of the output $\langle 2, 4 \rangle$ in red. This is illustrated in Figure 5.10.

Exercise 5.3.1. *For the linear transformation $T\langle x_1, x_2 \rangle = \langle -x_2, x_1 \rangle$ and each of the inputs given below, compute the corresponding output and draw a picture that contains both the input and the output, with inputs and outputs in different colors.*

1. $T(\langle 0, -3 \rangle)$
2. $T(\langle 3, 2 \rangle)$
3. $T(\langle -1, 1 \rangle)$
4. $T(\langle -2, 4 \rangle)$

The following two exercises (5.3.2 and 5.3.3) are designed to illustrate the two linearity properties (given in Definition 5.2.1) that make $T(\langle x_1, x_2 \rangle) = \langle -x_2, x_1 \rangle$ be a linear transformation.

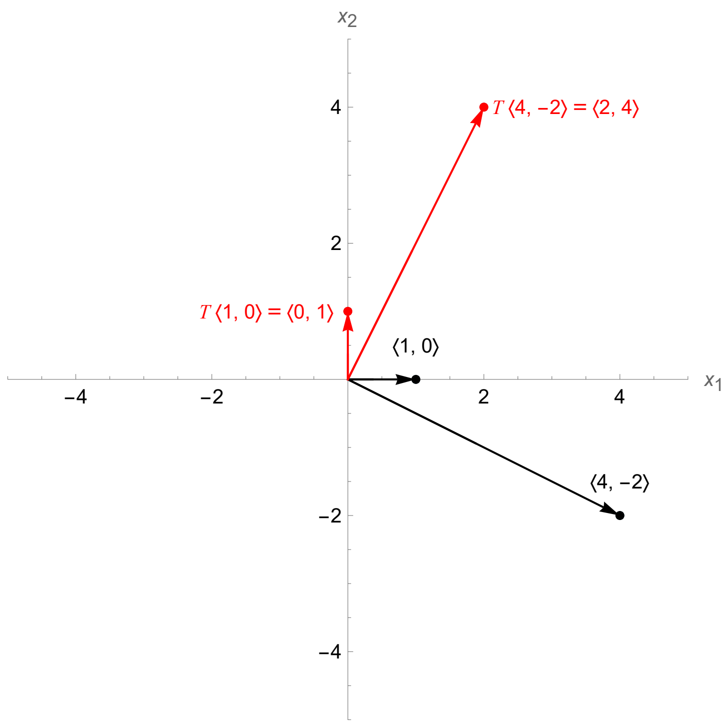


Figure 5.10: Mapping Two Vectors Using $T(\langle x_1, x_2\rangle) = (\langle -x_2, x_1\rangle)$

Exercise 5.3.2. *For the linear transformation $T\langle x_1, x_2\rangle = \langle -x_2, x_1\rangle$, draw a picture that contains all of the following:*

1. *the standard representatives of the vectors $\vec{x} = \langle 1, 0\rangle$ and and $\vec{y} = \langle 1, 1\rangle$.*

2. *the standard representative of the vector $\vec{x} + \vec{y}$ along with the other two sides of the parallelogram that you would use to illustrate the vector addition $\vec{x} + \vec{y}$ (as in Chapter 1).*

3. *the standard representatives of the vectors $T(\vec{x})$ and $T(\vec{y})$.*

4. *the standard representative of the vector $T(\vec{x}) + T(\vec{y})$ along with the other two sides of the parallelogram that you would use to illustrate the vector addition $T(\vec{x}) + T(\vec{y})$ (as in Chapter 1).*

5. *the standard representative of the vector $T(\vec{x} + \vec{y})$.*

If you did all of this correctly, you should see that

$$T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y}).$$

Exercise 5.3.3. *Let $\vec{x}$ be the vector $\vec{x} = \langle 1, -1\rangle$. For the linear transformation $T\langle x_1, x_2\rangle = \langle -x_2, x_1\rangle$, draw a picture that contains all of the following:*

1. *the standard representatives of the vectors $\vec{x}$ and $2\vec{x}$.*
2. *the standard representatives of the vectors $T(\vec{x})$ and $T(2\vec{x})$.*
3. *the standard representative of the vector $2T(\vec{x})$*

If you did all of this correctly, you should see that

$$T(2\vec{x}) = 2T(\vec{x}).$$

Perhaps by working Exercises 5.3.1, 5.3.2 and 5.3.3, you have started to get a sense of what the linear transformation $T(\langle x_1, x_2\rangle) = \langle -x_2, x_1\rangle$ "does" to its inputs. Can we describe in words what T does? In other words, can we describe in words how T "transforms" vectors in R^2? By working through Exercises 5.3.1, 5.3.2 and 5.3.3, you may have observed that the effect of applying T to a vector in R^2 is to rotate that vector counterclockwise by an angle of 90°. This is in fact what T does. We call this the **action** of T. Here is how we describe in words what the action of T is:

> The action of the linear transformation $T(\langle x_1, x_2\rangle) = \langle -x_2, x_1\rangle$ is to rotate nonzero vectors in R^2 though an angle of 90° counterclockwise.

Notice that in the above statement, we have added the caveat that T rotates all **nonzero** vectors by 90°. This is because

$$T(\langle 0, 0\rangle) = \langle 0, 0\rangle$$

and the concept of angle does not apply to the zero vector because the zero vector has no length or direction. The fact that $T(\langle 0, 0\rangle) = \langle 0, 0\rangle$ comes as no surprise because Theorem 5.2.2 tells us that any linear transformation maps zero vectors to zero vectors.

If you are a little uncomfortable that we have jumped to the conclusion that T rotates **all** nonzero vectors in R^2 based on just looking at some specific examples of these rotations in Exercises 5.3.2 and 5.3.3, then that is good! Looking at specific examples is not good enough to allow us to jump to a general claim such as the one we have made. Nonetheless, we suspect that the claim is correct and, fortunately, we can provide a mathematical justification of the claim.

To verify that the action of T is to rotate all non–zero vectors in R^2 counterclockwise by $90°$, first observe that if $\vec{x} = \langle x_1, x_2 \rangle$ is any nonzero vector in R^2, then the dot product of $\vec{x}$ with $T(\vec{x})$ is

$$\vec{x} \cdot T(\vec{x}) = \langle x_1, x_2 \rangle \cdot \langle -x_2, x_1 \rangle = (x_1)(-x_2) + (x_2)(x_1) = 0.$$

This tells us that $\vec{x}$ and $T(\vec{x})$ are orthogonal to each other (meaning that the angle between $\vec{x}$ and $T(\vec{x})$ is $90°$).

Secondly, note that $\vec{x}$ and $T(\vec{x})$ have the same magnitude because

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

and

$$\|T(\vec{x})\| = \sqrt{(-x_2)^2 + x_1^2} = \sqrt{x_1^2 + x_2^2}.$$

We have proved that T rotates each vector $\vec{x} \in R^2$ to a vector that is perpendicular to $\vec{x}$ and has the same length as $\vec{x}$. The only thing we have not proved is that the rotation is counterclockwise (rather than clockwise). We won't do that right here, but the direction of rotation will become evident when we study the general problem of rotation by and given angle θ in Section 5.3.6.

We have approached the problem of trying to visualize the action of $T(\langle x_1, x_2 \rangle) = \langle -x_2, x_1 \rangle$ in a rather informal way so far. We have just done a few sample calculations and drawn a few pictures. It is amazing that we have gotten so far in our understanding of T just by doing this! It seems that we now know pretty much everything there is to know about T. Now that we know that T rotates all nonzero vectors in R^2 counterclockwise through an angle of $90°$, we can answer the following questions just based on the mental picture of T that we have:

Is T invertible? Yes. If we choose any vector $\vec{y}$ in R^2 and rotate it $90°$ **clockwise**, then we obtain a unique vector $\vec{x}$ such that $T(\vec{x}) = \vec{y}$. In other words, there is a unique way to "undo" the action of T.

What is $\ker(T)$**?** $\ker(T) = \left\{\vec{0}_2\right\}$. If we take any nonzero vector and rotate it by 90° then we get a nonzero vector. Thus the only vector $\vec{x}$ such that $T(\vec{x}) = \vec{0}_2$ is $\vec{x} = \vec{0}_2$.

Let's us now check that the observations we have been able to make based on our informal investigation agree with the theory that we studied in Section 5.2.

First note that we can use Theorem 5.2.1 to find the matrix of T. Since

$$T(\vec{e}_1) = T(\langle 1, 0\rangle) = \langle 0, 1\rangle$$
$$T(\vec{e}_2) = T(\langle 0, 1\rangle) = \langle -1, 0\rangle,$$

then the matrix of T is

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

Since $\operatorname{rref}(A) = I_2$, then we can immediately see that T is invertible, that $\operatorname{Range}(T) = R^2$ and that $\ker(T) = \left\{\vec{0}_2\right\}$. Also, the inverse of the matrix A is

$$A^{-1} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

and this tells us that T^{-1} has formula

$$T^{-1}(x) = A^{-1}\vec{x} = \langle x_2, -x_1\rangle.$$

Exercise 5.3.4. *We have seen above that the inverse of* $T(\langle x_1, x_2\rangle) = \langle -x_2, x_1\rangle$ *is* $T^{-1}(\langle x_1, x_2\rangle) = \langle x_2, -x_1\rangle$*. Describe in words what the action of* T^{-1} *is and draw a few pictures to check your claim.*

We have undertaken quite a detailed examination of the linear transformation that rotates vectors in R^2 by 90° counterclockwise. We have purposely spent a lot of time on this example because it is our first illustration of what tools can be applied in trying to understand how linear transformations behave. In the examples we will study in Sections 5.3.2 - 5.3.5, we will not provide as much detail as we did with the current example. Instead, we will ask you to study the details in exercises. In Section 5.3.6, we will generalize the rotation transformation that we presented in the current section by considering rotations by any given angle θ.

The basic procedure for trying to understand the action of a linear transformation $T : R^2 \to R^2$ is:

1. Draw pictures! Pick out a few sample vectors and plot the inputs and their corresponding outputs. It is a good idea to draw things to scale using either graph paper and a ruler or some software such as *Desmos.*

2. Use the theory of Section 5.2 to determine information about invertibility, range and kernel.

3. Compare what you learned by drawing pictures and what you learned by using the theory. The pictures and the theory should agree with each other.

5.3.2 Example: A Projection Transformation

In this section we will study the linear transformation that *projects* every vector in R^2 onto the x_1 axis. The formula for this linear transformation is

$$T\left(\langle x_1, x_2\rangle\right) = \langle x_1, 0\rangle .$$

By doing a few computations and drawing a few pictures, we can quickly get the idea of the action of T. Every nonzero vector in R^2 is assigned to the vector that is the shadow that we would see on the x_1 axis if we were to draw the picture of $\vec{x}$ in standard position and then shine a light perpendicular to the x_1 axis. The action is illustrated in Figure 5.11

Exercise 5.3.5. *Draw pictures of the inputs $\langle -3, 5\rangle$ and $\langle 4, -4\rangle$ and their corresponding outputs under the projection transformation $T\left(\langle x_1, x_2\rangle\right) = \langle x_1, 0\rangle$.*

Exercise 5.3.6. *Just based on your picture drawing, try to answer the following questions.*

1. *What is the range of T? Give your answer in the form* $\text{Range}\,(T) = \text{Span}\,\{\_\_\_\_\_\_\_\}$.

2. *Does T map R^2 onto R^2?*

3. *Is T one–to–one? Hint: Does the equation $T\left(\vec{x}\right) = \langle 8, 0\rangle$ have a unique solution?*

4. *Is T invertible? Why or why not?*

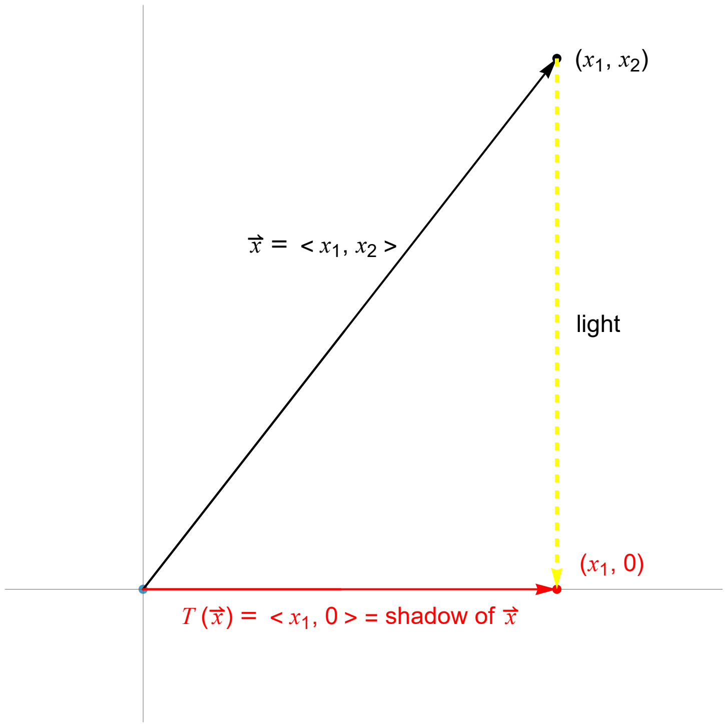


Figure 5.11: Projection of $\vec{x}$ onto the x_1 Axis

5. *What is the kernel of T? Give your answer in the form* $\ker(T) = \text{Span}\{\_\_\_\_\_\_\_\}$.

6. $\dim(\text{Range}(T)) = \_\_\_\_\_$ *and* $\dim(\ker(T)) = \_\_\_\_\_\_$, *and thus*

$$\dim(\text{Range}(T)) + \dim(\ker(T)) = \_\_\_\_\_\_.$$

Exercise 5.3.7. *Hopefully you were able to answer all of the questions in Exercise 5.3.6 by just drawing pictures. Now find the standard matrix, A, of T and use it to answer all of the same questions asked in Exercise 5.3.6.*

5.3.3 Example: Scaling Transformations

Consider the linear transformation $T : R^2 \to R^2$ defined by

$$T(\vec{x}) = 2\vec{x}.$$

If $\vec{x} = \langle x_1, x_2 \rangle$, then $2\vec{x} = \langle 2x_1, 2x_2 \rangle$, so we can write the formula for T as

$$T(\langle x_1, x_2 \rangle) = \langle 2x_1, 2x_2 \rangle,$$

but for the purpose of answering some questions (especially in trying to form a general picture in our minds of the action of T) it is convenient to just think of T as $T(\vec{x}) = 2\vec{x}$.

Exercise 5.3.8. *For the linear transformation $T(\vec{x}) = 2\vec{x}$,*

1. *Plot the inputs $\langle 3, 4 \rangle$ and $\langle -2, 2 \rangle$ along with their corresponding outputs under T.*
2. *We know from what we learned in Chapter 1 that if $\vec{x}$ is any nonzero vector, then the magnitude of the vector $2\vec{x}$ is _____ times the magnitude of $\vec{x}$ and that the vector $2\vec{x}$ points in the ______ direction that $\vec{x}$ points.*

Exercise 5.3.9. *Based on your drawing of pictures and on your answers in Exercise 5.3.8, try to answer the following questions concerning the linear transformation $T(\vec{x}) = 2\vec{x}$.*

1. *What is the range of T? Give your answer in the form* $\text{Range}(T) = \text{Span}\{\_\_\_\_\_\_\_\}$.
2. *Is T invertible? If so, find a formula for T^{-1}.*
3. *What is the kernel of T? Give your answer in the form* $\ker(T) = \text{Span}\{\_\_\_\_\_\_\_\}$.
4. $\dim(\text{Range}(T)) = \_\_\_\_\_$ *and* $\dim(\ker(T)) = \_\_\_\_\_\_$, *and thus*

$$\dim(\text{Range}(T)) + \dim(\ker(T)) = \_\_\_\_\_\_.$$

Exercise 5.3.10. *Write down the standard matrix, A, of the linear transformation $T(\vec{x}) = 2\vec{x}$ and use it to answer all of the same questions that were asked about T in the previous exercise.*

Linear transformations of the form $T(\vec{x}) = r\vec{x}$, where r is a scalar are called scaling transformations. If $r > 1$ then the transformation is called a **dilation** or **stretching**. If $0 < r < 1$ then the transformation is called a **contraction** or **shrinking**.

Exercise 5.3.11. *What is the linear transformation, T, from R^2 to R^2 such that for all nonzero $\vec{x} \in R^2$, $T(\vec{x})$ has 4 times the magnitude of $\vec{x}$ and points in the opposite direction of x.*

Exercise 5.3.12. *What is the linear transformation, T, from R^2 to R^2 such that for all nonzero $\vec{x} \in R^2$, $T(\vec{x})$ has 1/3 the magnitude of $\vec{x}$ and points in the same direction of x.*

5.3.4 Example: A Reflection Transformation

A reflection transformation is one that reflects every vector in R^2 across some given line (which passes through the origin) in R^2. The reflection of a vector $\vec{x}$ through the line L is the vector $T(\vec{x})$ such that $T(\vec{x})$ is the mirror image of $\vec{x}$ using the line L as the mirror. In this example we will study the linear transformation that reflects vectors through the mirror which is the line $x_2 = x_1$. This reflection is illustrated in Figure 5.12 for the vector $\vec{x} = \langle 2, 4 \rangle$.

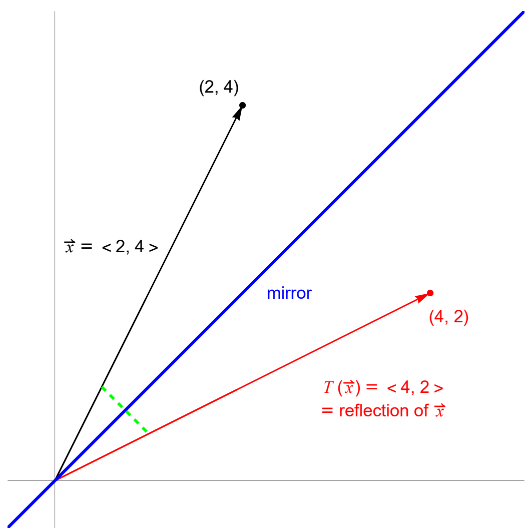


Figure 5.12: Reflection through the $x_2 = x_1$ Mirror

Exercise 5.3.13. *Let $T : R^2 \to R^2$ be the reflection transformation described above. Find $T(\langle 1, 0 \rangle)$ and $T(\langle 0, 1 \rangle)$. Based on this, using Theorem 5.2.1, you can write down the standard matrix, A, of T, and then you can write a formula for T in the form $T(\langle x_1, x_2 \rangle) =$ ______.*

Exercise 5.3.14. *Answer the following questions concerning the reflection linear transformation.*

1. *What is the range of T? Give your answer in the form* $\text{Range}(T) = \text{Span}\{\_\_\_\_\_\_\_\}$.

2. *Is T invertible? If so, find a formula for T^{-1}.*

3. *What is the kernel of T? Give your answer in the form* $\ker(T) = \text{Span}\{\_\_\_\_\_\_\_\}$.

4. $\dim(\text{Range}(T)) = \_\_\_\_\_$ *and* $\dim(\ker(T)) = \_\_\_\_\_\_$, *and thus*

$$\dim(\text{Range}(T)) + \dim(\ker(T)) = \_\_\_\_\_\_.$$

Exercise 5.3.15. *Use the formula for T that you found to show that for any $\vec{x} \in R^2$ it is true that $\vec{x}$ and $T(\vec{x})$ have the same magnitude. Does this make sense based on pictures?*

5.3.5 Example: A Shearing Transformation

As our final example of visualizing linear transformations, we will consider what is known as a shearing transformation. We will consider the shearing transformation $T : R^2 \to R^2$ defined by

$$T(\langle x_1, x_2\rangle) = \langle x_1 + x_2, x_2\rangle .$$

We can understand the action of T by strategically choosing some inputs and plotting them together with their outputs. It requires a little more plotting than the amount that was needed in the previous examples we looked at to see what the shearing does.

Exercise 5.3.16. *Plot the inputs*

$$\begin{array}{ccc} \langle -1, 2\rangle & \langle 0, 2\rangle & \langle 1, 2\rangle \\ \langle -1, 1\rangle & \langle 0, 1\rangle & \langle 1, 1\rangle \\ \langle -1, 0\rangle & \langle 0, 0\rangle & \langle 1, 0\rangle \\ \langle -1, -1\rangle & \langle 0, -1\rangle & \langle 1, -1\rangle \\ \langle -1, -2\rangle & \langle 0, -2\rangle & \langle 1, -2\rangle \end{array}$$

and their corresponding outputs. Plot as many inputs and outputs as you need to in order to get a visual picture of the action of T. Then write a sentence of two that describes the action of T.

Exercise 5.3.17. *For the linear transformation* $T(\langle x_1, x_2 \rangle) = \langle x_1 + x_2, x_2 \rangle$*, find the standard matrix,* A*, of* T*.*

Exercise 5.3.18. *Answer the following questions regarding the shearing transformation* $T(\langle x_1, x_2 \rangle) = \langle x_1 + x_2, x_2 \rangle$*.*

1. *What is the range of* T*? Give your answer in the form* $\text{Range}(T) = \text{Span}\{\_\_\_\_\_\_\_\}$.
2. *Is* T *invertible? If so, find a formula for* T^{-1}*.*
3. *What is the kernel of* T*? Give your answer in the form* $\ker(T) = \text{Span}\{\_\_\_\_\_\_\_\}$.
4. $\dim(\text{Range}(T)) = \_\_\_\_\_$ *and* $\dim(\ker(T)) = \_\_\_\_\_\_$*, and thus*

$$\dim(\text{Range}(T)) + \dim(\ker(T)) = \_\_\_\_\_\_.$$

5.3.6 Rotation Transformations in General

In Section 5.3.1, we carried out a careful investigation of the linear transformation that rotates all vectors in R^2 by 90° counterclockwise. That linear transformation is defined by

$$T(\langle x_1, x_2 \rangle) = \langle -x_2, x_1 \rangle .$$

Its matrix is

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

We will now investigate how rotations work in general. For a given angle θ, we will find the linear transformation that rotates all vectors in R^2 through an angle of θ. If $\theta > 0$, then we get a counterclockwise rotation. If $\theta < 0$, then we get a clockwise rotation. Since we are studying a whole class of linear transformations, we should use some appropriate notation to name these transformations. We don't want to just use the name T over and over again to refer to different linear transformations that are all related to one another. We will use the name R_θ to name the linear transformation that rotates all vectors in R^2 by angle θ. Thus, for example, R_{90° is the transformation that rotates all vectors by 90° counterclockwise and R_{-30° is the transformation that rotates all vectors by 30° clockwise. Since the inverse of rotating by angle θ is to rotate by $-\theta$, then

$$R_\theta^{-1} = R_{-\theta}.$$

We will be finding the matrices for the linear transformations R_θ and we will name these matrices A_θ.

To get familiar with what we are going to do, including the notation described above, let's first look at a specific example. We will then move on to the general treatment.

Example 5.3.1. *We will find the linear transformation, R_{45°, that rotates all vectors in R^2 through an angle of 45° counterclockwise.*

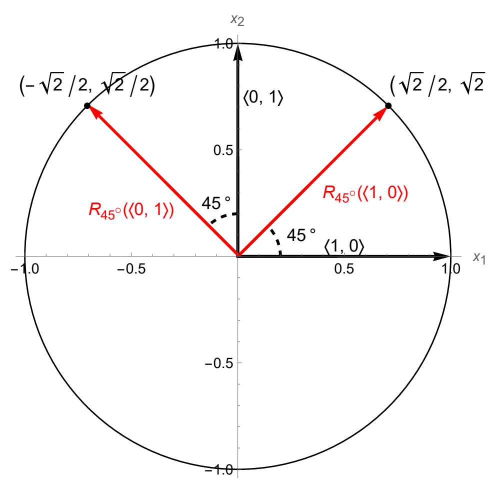


Figure 5.13: R_{45° = Rotation by 45° Counterclockwise

By using our knowledge of the unit circle and looking at Figure 5.13, we see that

$$R_{45^\circ}\left(\langle 1,0\rangle\right) = \left\langle \sqrt{2}/2, \sqrt{2}/2 \right\rangle$$
$$R_{45^\circ}\left(\langle 0,1\rangle\right) = \left\langle -\sqrt{2}/2, \sqrt{2}/2 \right\rangle.$$

Theorem 5.2.1 tell use that the matrix for R_{45° is

$$A_{45^\circ} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}.$$

Thus $R_{45^\circ}(\langle x_1, x_2\rangle) = A_{45^\circ}\langle x_1, x_2\rangle$ *for all* $\langle x_1, x_2\rangle$ *in* R^2*. If we want to, we can write the formula for* R_{45° *as*

$$R_{45^\circ}(\langle x_1, x_2\rangle) = \left\langle \frac{\sqrt{2}}{2}x_1 - \frac{\sqrt{2}}{2}x_2, \frac{\sqrt{2}}{2}x_1 + \frac{\sqrt{2}}{2}x_2 \right\rangle.$$

R_{45° *is invertible and its inverse is* $R_{45^\circ}^{-1} = R_{-45^\circ}$*. The matrix for* $R_{45^\circ}^{-1}$ *is*

$$A_{45^\circ}^{-1} = A_{-45^\circ} = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}.$$

Exercise 5.3.19. *In Example 5.3.1 we found the linear transformation* R_{45° *that rotates all angles in* R^2 *by an angle of* 45° *counterclockwise. We also found* $R_{45^\circ}^{-1} = R_{-45^\circ}$*. Try to come up with the answers to the following without using any of the formulas found in Example 5.3.1. Just do it based on your knowledge of the unit circle. Then use the formulas to see if you got it right.*

1. $R_{45^\circ}\left(\left\langle \sqrt{2}/2, \sqrt{2}/2\right\rangle\right)$
2. $R_{45^\circ}\left(\left\langle \sqrt{2}/2, -\sqrt{2}/2\right\rangle\right)$
3. $R_{45^\circ}(\langle -1, 0\rangle)$
4. $R_{45^\circ}^{-1}(\langle 0, -1\rangle)$.

Exercise 5.3.20. *Find the linear transformation,* R_{30°*, that rotates all vectors in* R^2 *through an angle of* 30° *counterclockwise. Then find* $R_{30^\circ}^{-1}$*.*

To find the linear transformation R_θ for a general θ, we draw the general unit circle picture shown in Figure 5.14.

From the Figure, we see that

$$\begin{aligned} R_\theta(\langle 1, 0\rangle) &= \langle \cos(\theta), \sin(\theta)\rangle \\ R_\theta(\langle 0, 1\rangle) &= \langle \cos(90^\circ + \theta), \sin(90^\circ + \theta)\rangle. \end{aligned}$$

We now use the trigonometric identities

$$\begin{aligned} \cos(90^\circ + \theta) &= -\sin(\theta) \\ \sin(90^\circ + \theta) &= \cos(\theta) \end{aligned}$$

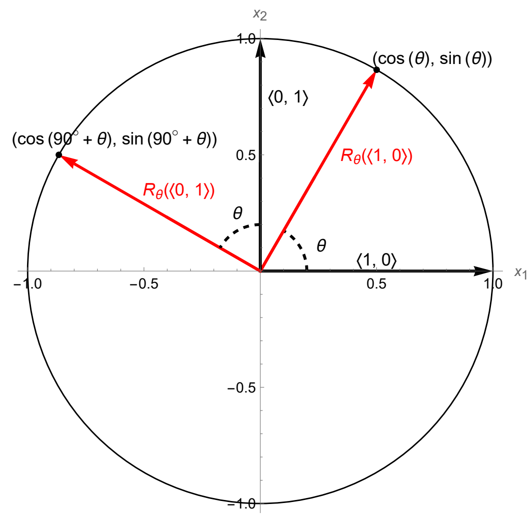


Figure 5.14: $R_\theta =$ Rotation by Angle θ

to obtain

$$R_\theta\left(\langle 0,1\rangle\right) = \langle -\sin\left(\theta\right), \cos\left(\theta\right)\rangle .$$

Hence the matrix for R_θ is

$$A_\theta = \begin{bmatrix} \cos\left(\theta\right) & -\sin\left(\theta\right) \\ \sin\left(\theta\right) & \cos\left(\theta\right) \end{bmatrix} . \tag{5.12}$$

Let's plug a few specific θ values into formula (5.12). If we plug in $\theta = 90°$, we get

$$A_\theta = \begin{bmatrix} \cos\left(90°\right) & -\sin\left(90°\right) \\ \sin\left(90°\right) & \cos\left(90°\right) \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

which is the same thing we got in Section 5.3.1. If we plug in $\theta = 45°$, we get

$$A_\theta = \begin{bmatrix} \cos\left(45°\right) & -\sin\left(45°\right) \\ \sin\left(45°\right) & \cos\left(45°\right) \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$$

which is the same thing we got in Example 5.3.1.

We have now found the linear transformation that rotates all vectors in R^2 by a given angle θ. It is $R_\theta\left(\vec{x}\right) = A_\theta \vec{x}$ where A_θ is the matrix given by formula (5.12). We can also write the formula for R_θ as

$$R_\theta\left(\langle x_1, x_2\rangle\right) = \langle \cos\left(\theta\right) x_1 - \sin\left(\theta\right) x_2, \sin\left(\theta\right) x_1 + \cos\left(\theta\right) x_2\rangle . \tag{5.13}$$

Exercise 5.3.21. *For any given angle θ, the matrix, A_θ, for R_θ is given by formula (5.12). Find the matrix, $A_\theta^{-1} = A_{-\theta}$, for $R_\theta^{-1} = R_{-\theta}$.*

Hint: Just plug in $-\theta$ in place of θ in formula (5.12) and use the even–odd identities for sine and cosine which are

$$\cos(-\theta) = \cos(\theta)$$
$$\sin(-\theta) = -\sin(\theta) .$$

Exercise 5.3.22. *In Exercise 5.3.21, you found A_θ^{-1}. Check that you got the right answer by computing $A_\theta^{-1} A_\theta$.*

Exercise 5.3.23. *Since the linear transformation R_θ rotates all vectors in R^2 by angle θ, it should be true for any $\vec{x}$ that the magnitude of $R_\theta(\vec{x})$ is the same as the magnitude of $\vec{x}$. Verify that this is true by using the formula (5.13).*

Exercise 5.3.24. *Use the formula (5.13) to verify that if $\vec{x} = \langle x_1, x_2 \rangle$ is any nonzero vector in R^2 and θ is any angle, then*

$$\vec{x} \cdot R_\theta(\vec{x}) = \|\vec{x}\|^2 \cos(\theta) .$$

Conclude that $\vec{x} \cdot R_\theta(\vec{x}) > 0$ if θ is acute and that $\vec{x} \cdot R_\theta(\vec{x}) < 0$ if θ is obtuse. What is true about $\vec{x} \cdot R_\theta(\vec{x})$ when $\theta = 90^\circ$ or when $\vec{x} = \vec{0}_2$?

Exercise 5.3.25. *Suppose that α and θ are some given angles. The point $(\cos(\alpha), \sin(\alpha))$ is the point on the unit circle shown in Figure 5.15.*

If we rotate the vector $\vec{x} = \langle \cos(\alpha), \sin(\alpha) \rangle$ through angle θ by applying the linear transformation R_θ, then we get the vector

$$R_\theta(\vec{x}) = \langle \cos(\alpha+\theta), \sin(\alpha+\theta) \rangle . \tag{5.14}$$

The above equation is true for any angles α and θ.

Use the formula (5.13) to compute $R_\theta(\vec{x})$ and compare what you get to the formula (5.14) for $R_\theta(\vec{x})$ shown above. You should see two familiar trigonometric identities pop out.

Repeat this exercise by computing $R_{-\theta}(\vec{x})$ using both formula (5.14) and formula (5.13). You should obtain two more familiar trigonometric identities.

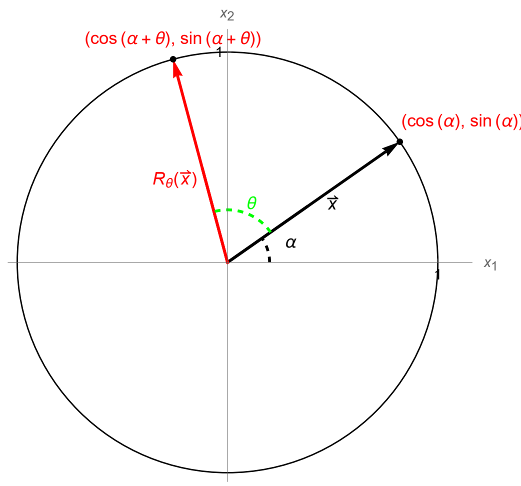


Figure 5.15: Rotation by Angle θ

5.4 Linear Transformation of Lines

Recall that at the beginning of this chapter we said that linear transformations are functions that map lines to lines or points. In Section 5.2, we provided our formal definition of what a linear transformation is. Definition 5.2.1 tells us that a linear transformation $T : R^n \to R^m$ is a function that satisfies the linearity properties

$$T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$$
$$T(c\vec{x}) = cT(\vec{x})$$

for all vectors $\vec{x}$ and $\vec{y}$ in R^n and for all scalars c. It was mentioned right after we stated Definition 5.2.1 that it might not be clear to you right after reading the definition that the definition describes a class of functions that map lines to lines or points. In Section 5.3, you had ample opportunity to acquaint yourself with some specific linear transformations from R^2 to R^2 and hopefully that helped you to start seeing that linear transformations do map lines to lines or points. In this section we will verify that this is true.

Before we can show that linear transformations map lines to lines or points, we need to be sure that we have a clear understanding of what we mean by a "line". In previous math courses you have taken, you learned how to draw pictures of lines in R^2 and write equations for these lines. One way to determine a line, L, in R^2 is to be given two points, $P = (p_1, p_2)$ and

$Q = (q_1, q_2)$ on the line. Using those two points (and assuming that $q_1 \neq p_1$, which means that L is not a vertical line), you can compute the slope,

$$m = \frac{q_2 - p_2}{q_1 - p_1},$$

of the line and then write an equation for the line using the point–slope equation

$$x_2 - p_2 = m\,(x_1 - p_1)\,.$$

As an example, let us find an equation for the line, L, in R^2 that contains the points $P = (0, 2)$ and $Q = (4, -5)$. To do this, we compute the slope,

$$m = \frac{-5 - 2}{4 - 0} = -\frac{7}{4}$$

and then we see that the line has equation

$$L : x_2 - 2 = -\frac{7}{4}\,(x_1 - 0)$$

which can be written as

$$L : x_2 = -\frac{7}{4}x_1 + 2.$$

The point–slope method for writing the equation of a line in R^2 can always be used, except when L is a vertical line. This is because the slope of a vertical line is undefined. If L is a vertical line that contains the point $P = (p_1, p_2)$, then the equation of L is

$$L : x_1 = p_1.$$

For example, the equation of the vertical line that contains the point $P = (3, 2)$ is

$$L : x_1 = 3.$$

Exercise 5.4.1. *Write equations for the lines, L, that are described below. Draw a picture of each line.*

1. *L contains the points $P = (3, 1)$ and $Q = (2, -4)$.*
2. *L contains the points $P = (-4, -1)$ and $Q = (-3, -3)$.*

3. *L contains the points $P = (3, 5)$ and $Q = (-12, 5)$.*

4. *L contains the points $P = (3, -2)$ and $Q = (3, -1)$.*

Although the point–slope method is certainly nice for writing equations of lines in R^2, there is no way to adapt this approach to write equations for lines in R^3. That is because there is no way to compute a single number, m, that could be considered the "slope" of a line in R^3. Think about it: Suppose you are given the two points $P = (-1, 2, 3)$ and $Q = (-1, 1, -1)$ in R^3. There is a unique line, L, in R^3 that contains these two points. But what is its slope? We can't really define it. However, what we can define is a direction vector for L. A direction vector for L is any vector that is parallel to L. If we are given a point, P, on L, and a direction vector, $\vec{d}$, for L, then we can write a set of equations that describe L. We can also write a single (vector form) equation for L. This leads us to our formal definition of what we mean by a line in R^n (for any $n \geq 2$).

Definition 5.4.1. *Suppose that $P = (p_1, p_2, \ldots, p_n)$ is a point in R^n and suppose that $\vec{d} = \langle d_1, d_2, \ldots, d_n \rangle$ is a non–zero vector in R^n. The **line**, L, that contains the point P and has **direction vector** $\vec{d}$ is the set of all points $X = (x_1, x_2, \ldots, x_n)$ such that*

1. *The point P lies on L and*

2. *L is parallel to $\vec{d}$.*

We can form a visual picture of lines in R^2 and in R^3. In R^n, when $n > 3$, visualizing a line is difficult but the mathematics of working with lines in R^n works the same for any n. Figure 5.16 illustrates this idea in R^2.

In the Figure, we have drawn the line L that contains the point $P = (p_1, p_2)$ and is parallel to the non–zero vector $\vec{d} = \langle d_1, d_2 \rangle$. If $X = (x_1, x_2)$ is any arbitrarily chosen point on L (other than the point P itself), then the vector

$$\overrightarrow{PX} = \langle x_1 - p_1, x_2 - p_2 \rangle$$

is parallel to the vector $\vec{d} = \langle d_1, d_2 \rangle$. This means that $\overrightarrow{PX}$ is a scalar multiple of $\vec{d}$ and hence there is some scalar t such that $\overrightarrow{PX} = t\vec{d}$. Writing this out in detail, we obtain

$$\langle x_1 - p_1, x_2 - p_2 \rangle = t \langle d_1, d_2 \rangle . \tag{5.15}$$

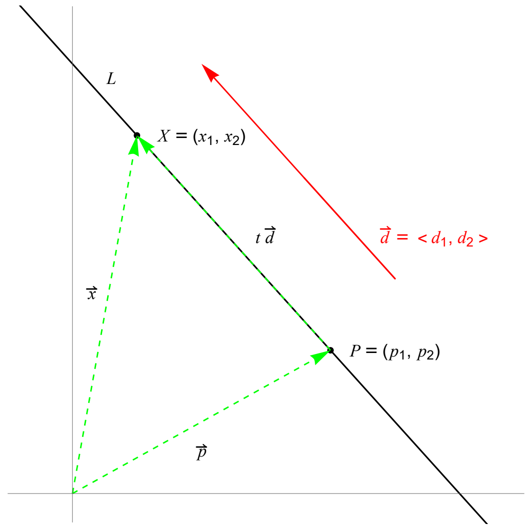


Figure 5.16: $\vec{x} = \vec{p} + t\vec{d}$

Equation (5.15) tells us that

$$\begin{aligned} x_1 - p_1 &= td_1 \\ x_2 - p_2 &= td_2 \end{aligned}$$

and hence

$$\begin{aligned} x_1 &= p_1 + td_1 \\ x_2 &= p_2 + td_2 \\ -\infty &< t < \infty \end{aligned} \tag{5.16}$$

The equations (5.16) are what we refer to as **parametric equations** of the line, L, that contains the point $P = (p_1, p_2)$ and has direction vector $\vec{d} = \langle d_1, d_2 \rangle$. When we write $-\infty < t < \infty$, we are saying that we are allowing the scalar t (which is also referred to as a **parameter**) to range over all real numbers. This traces out the entire line L. The line is infinitely long. If we wish to only consider some piece of L, i.e., a line segment, then we can restrict the parameter t to only be allowed to range over some smaller interval. For example, instead of writing $-\infty < t < \infty$ we could write $0 \le t \le 1$ and that would trace out only some line segment that lies on L.

Another way to interpret equation (5.15) is in terms of vectors. We can write equation (5.15) as

$$\langle x_1, x_2 \rangle - \langle p_1, p_2 \rangle = t \langle d_1, d_2 \rangle$$

or

$$\langle x_1, x_2 \rangle = \langle p_1, p_2 \rangle + t \langle d_1, d_2 \rangle .$$

If we define $\vec{x} = OX = \langle x_1, x_2 \rangle$ and $\vec{p} = \overrightarrow{OP} = \langle p_1, p_2 \rangle$ (as in Figure 5.16), then we can write the above equation as

$$\begin{gathered} \vec{x} = \vec{p} + t\vec{d} \\ -\infty < t < \infty. \end{gathered} \tag{5.17}$$

Equation (5.17) is called a **vector parametric equation** (or, more briefly, a **vector equation**) of the line L. Again, we have included $-\infty < t < \infty$ in the vector parametric description above because this produces the entire line L but we could also produce a line segment by being more restrictive with the parameter t. Moving forward, when we write parametric equations or vector equations for lines, we will sometimes choose to save space by not writing $-\infty < t < \infty$. When we do this, it will be understood that we are considering the entire line and allowing $-\infty < t < \infty$. If we wish to consider only a line segment, then we will specify $a \leq t \leq b$, indicating that the parameter is being restricted to the interval $[a, b]$.

Example 5.4.1. *Let L be the line in R^2 that contains the point $P = (2, 1)$ and has direction vector $\vec{d} = \langle 2, -3 \rangle$. Draw a picture of this line and:*

1. *Write parametric equations for L.*

2. *Write a vector equation for L.*

Solution: *A picture of L is shown in Figure 5.17.*

If $X = (x_1, x_2)$ is any point on L, then $\overrightarrow{PX} = t\vec{d}$ for some scalar t and thus

$$\langle x_1 - 2, x_2 - 1 \rangle = t \langle 2, -3 \rangle .$$

From this we see that parametric equations for L are

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

and a vector equation for L is

$$\langle x_1, x_2 \rangle = \langle 2, 1 \rangle + t \langle 2, -3 \rangle .$$

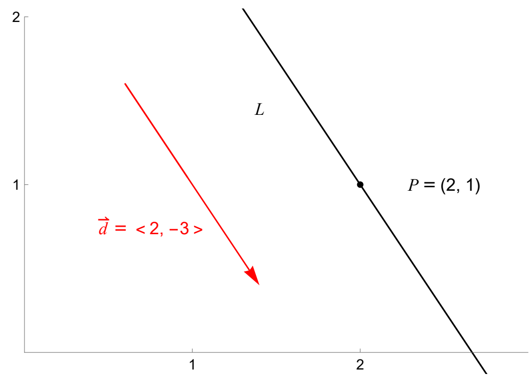


Figure 5.17: $\langle x_1, x_2\rangle = \langle 2, 1\rangle + t\langle 2, -3\rangle$

Example 5.4.2. *Write parametric equations and vector equations for the line, L, in R^3 that contains the points $P = (-1, 2, 3)$ and $Q = (-1, 1, -1)$.*

Solution: *The directed line segment from the point P to the point Q gives us a direction vector for L:*

$$\vec{d} = \overrightarrow{PQ} = \langle -1 - (-1), 1 - 2, -1 - 3\rangle = \langle 0, -1, -4\rangle .$$

A vector equation for L is

$$L : \vec{x} = \vec{p} + t\vec{d}$$

which is

$$L : \langle x_1, x_2, x_3\rangle = \langle -1, 2, 3\rangle + t \langle 0, -1, -4\rangle .$$

Parametric equations for L are

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

To check that this is correct, note that when we plug $t = 0$ into the parametric equations we get the point $P = (-1, 2, 3)$ and when we plug $t = 1$ into the parametric equations we get the point $Q = (-1, 1, -1)$.

Exercise 5.4.2. *For each of the lines, L, that are described below, write parametric equations and a vector equation for L. Graph the ones that are in R^2.*

1. *L is the line in R^2 that contains the point $P = (4, -2)$ and has direction vector $\vec{d} = \langle 0, -1 \rangle$.*

2. *L is the line in R^2 that contains the point $P = (3, 2)$ and has direction vector $\vec{d} = \langle 2, 2 \rangle$.*

3. *L is the line in R^2 that contains the points $P = (0, 2)$ and $Q = (-2, -2)$.*

4. *L is the line in R^2 that has equation $x_2 = 0$.*

5. *L is the line in R^2 that has equation $x_1 = 7$.*

6. *L is the line in R^2 that has equation $x_2 = x_1 + 4$.*

7. *L is the line in R^3 that contains the point $P = (2, -1, -1)$ and has direction vector $\vec{d} = \langle 1, 2, -2 \rangle$.*

8. *L is the line in R^3 that contains the point $P = (-1, 1, 1)$ and has direction vector $\vec{d} = \langle 0, 0, -1 \rangle$.*

9. *L is the line in R^3 that contains the two points $P = (-1, 2, 3)$ and $Q = (-1, 1, -1)$.*

10. *L is the line in R^4 that contains the two points $P = (4, -1, -2, -4)$ and $Q = (3, 2, -2, -2)$.*

We are now prepared to prove that any linear transformation from R^2 to R^2 maps any line to a line or to a point. Recall, from Section 5.1, the notation that we use for a function $f : D \to C$ to denote the image under f of a subset of the domain. If S is a subset of the domain, D, then $f(S)$ denotes the image of S under f. That is,

$$f(S) = \{f(x) \mid x \in S\}.$$

For a linear transformation $T : R^n \to R^m$, a line, L, in R^n is a subset of R^n, and we denote its image under T by $T(L)$.

Theorem 5.4.1. *Suppose that $T : R^n \to R^m$ is a linear transformation and suppose that L is a line in R^n. Specifically, suppose that*

$$L : \vec{x} = \vec{p} + t\vec{d} \tag{5.18}$$

where $\vec{d} \neq \vec{0}_n$.

Then $T(L)$ *is either a point or a line in* R^m. *Specifically,*

1. *If* $\vec{d} \notin \ker(T)$, *then* $T(L)$ *is a line in* R^m.

2. *If* $\vec{d} \in \ker(T)$, *then* $T(L)$ *is a point in* R^m.

Proof. Before we start the proof, let us review what equation (5.18) means:

$$\begin{aligned} P &= (p_1, p_2) \text{ is a certain point on } L \\ \vec{p} &= \langle p_1, p_2 \rangle = \overrightarrow{OP} \\ \vec{d} &\neq \vec{0}_n \text{ is a direction vector for } L \\ & t \text{ is a scalar variable} \\ X &= (x_1, x_2) \text{ is an arbitrary point on } L. \\ & X \text{ varies as } t \text{ varies.} \\ \vec{x} &= \langle x_1, x_2 \rangle = \overrightarrow{OX}. \end{aligned}$$

This is all summarized in Figure 5.16 (for the case of a line in R^2).

If we take any point X on L and apply the linear transformation T to $\vec{x}$, then we obtain (by using the linearity properties of T)

$$\vec{y} = T(\vec{x}) = T\left(\vec{p} + t\vec{d}\right) = T(\vec{p}) + tT\left(\vec{d}\right).$$

Hence

$$T(L) = \{T(\vec{x}) \mid \vec{x} \in L\} = \left\{T(\vec{p}) + tT\left(\vec{d}\right) \mid -\infty < t < \infty\right\}.$$

We see that if $\vec{d} \notin \ker(T)$, meaning that $T\left(\vec{d}\right) \neq \vec{0}_m$, then $T(L)$ is a line that contains the point $T(\vec{p})$ and has direction vector $T\left(\vec{d}\right)$. An equation for this line is

$$T(L) : \vec{y} = T(\vec{p}) + tT\left(\vec{d}\right).$$

If $\vec{d} \in \ker(T)$, then $T\left(\vec{d}\right) = \vec{0}_m$ and we simply have $T(L) = \{T(\vec{p})\}$, meaning that $T(L)$ consists of a single point. □

A nice thing about Theorem 5.4.1 is that it not only tells us that linear transformations map lines to lines or points, but it also tells us the criterion that determines whether a particular line is mapped to a line or a point. We will illustrate using a few examples from the linear transformations that we studied in Section 5.3.

Example 5.4.3. *In Section 5.3.1, we studied the linear transformation* $T : R^2 \to R^2$ *defined by*

$$T(\langle x_1, x_2\rangle) = (\langle -x_2, x_1\rangle)$$

and we discovered that this transformation rotates all vectors in R^2 *by an angle of* $90°$ *counterclockwise without changing the lengths of the vectors. This linear transformation is invertible meaning that*

$$\ker(T) = \left\{\vec{0}_2\right\}.$$

Hence if

$$L : \vec{x} = \vec{p} + t\vec{d}$$

is any line in R^2*, then since* $\vec{d} \neq \vec{0}_2$ *and* $\ker(T)$ *contains only the vector* $\vec{0}_2$*, we see that* $\vec{d} \notin \ker(T)$*. Therefore* $T(L)$ *is a line. This linear transformation maps all lines in* R^2 *to lines in* R^2*.*

If we are given a particular line, L*, in* R^2*, then we can easily find the line* $T(L)$*. For example, suppose that* L *is the horizontal line*

$$L : \ x_2 = 8$$

that is pictured in Figure 5.18. Every point on L *has the form* $(t, 8)$ *where* t *can be any value. Now note that*

$$T(\langle t, 8\rangle) = (\langle -8, t\rangle).$$

The points of the form $(-8, t)$ *lie on the vertical line* $x_1 = -8$*. Thus every point on the horizontal line* $x_2 = 8$ *is a mapped by* T *to a point on the vertical line* $x_1 = -8$ *as shown in Figure 5.18.*

Exercise 5.4.3. *Let* T *be the rotation transformation*

$$T(\langle x_1, x_2\rangle) = (\langle -x_2, x_1\rangle)$$

and let L *be the line*

$$L : \ x_2 = 2x_1 + 1.$$

Find $T(L)$ *and graph* L *and* $T(L)$ *together.*

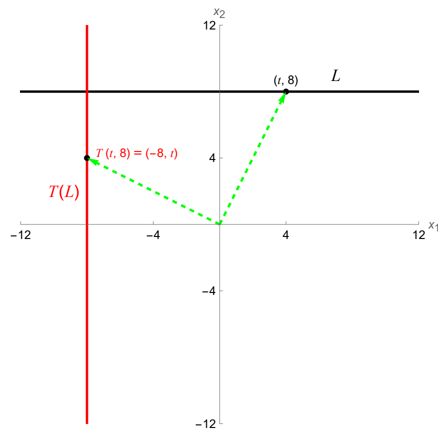


Figure 5.18: Line $x_2 = 8$ Mapped to Line $x_1 = -8$

Example 5.4.4. *Figure 5.19 shows a picture of the capital letter "A" which was constructed by choosing five points in R^2 and then connecting these points with line segments to form the letter "A".*

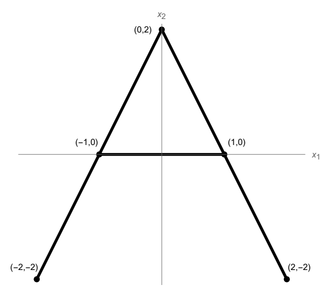


Figure 5.19: Letter "A"

Figure 5.20 shows the transformation of the letter "A" that is obtained by applying the linear transformation $T(\langle x_1, x_2 \rangle) = \langle -x_2, x_1 \rangle$. As can be seen, the transformation rotates the "A" counterclockwise by $90°$. Each of the three line segments that make up the original "A" are mapped to line segments on the transformed "A".

Example 5.4.5. *In Section 5.3.2 we studied the linear transformation*

$$T(\langle x_1, x_2 \rangle) = \langle x_1, 0 \rangle$$

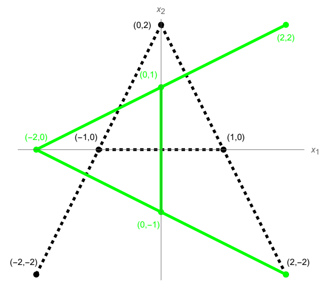


Figure 5.20: Letter "A" Rotated 90°

which projects all vectors in $\vec{x} \in R^2$ onto the x_1 axis. Since all vectors are projected onto the x_1 axis, then

$$\text{Range}(T) = \text{Span}\{\langle 1, 0\rangle\}.$$

The kernel of T consists of all solutions of the equation

$$T(\langle x_1, x_2\rangle) = \langle 0, 0\rangle,$$

which can be written, using the formula for T, as

$$\langle x_1, 0\rangle = \langle 0, 0\rangle.$$

Any vector of the form $\langle x_1, x_2\rangle = \langle 0, t\rangle$, where t can be any real number, satisfies the above equation and thus

$$\ker(T) = \text{Span}\{\langle 0, 1\rangle\}.$$

According to Theorem 5.4.1, any line in R^2 that does not have direction vector $\vec{d} = \langle 0, 1\rangle$ is mapped to a line. (In this case, the line is the x_1 axis.) However, if a line does have direction vector $\vec{d} = \langle 0, 1\rangle$, then that line is mapped to a point (which is a single point on the x_1 axis). The only lines that have direction vector $\vec{d} = \langle 0, 1\rangle$ are vertical lines. Hence,

- *If L is not a vertical line, then $T(L) =$ the x_1 axis*
- *If $L : \langle k, 0\rangle + t\langle 0, 1\rangle$ is a vertical line, then $T(L) = \{\langle k, 0\rangle\}$.*

Exercise 5.4.4. *In Section 5.3.3, we studied the linear transformation*

$$T(\vec{x}) = 2\vec{x}$$

which maps all vectors in R^2 to a vector pointing in the same direction but having twice the length.

1. *Explain why there are no lines in R^2 that are mapped by T to a point. (Hint: What is* $\ker(T)$*?)*

2. *Show that every line, L, in R^2 is mapped by T to a line that is parallel to L. In other words, show that if L is any line in R^2, then $T(L)$ is parallel to L.*

3. *Show that if L is a line that contains the point $(0,0)$, then $T(L) = L$.*

Exercise 5.4.5. *In Section 5.3.4 we studied the linear transformation*

$$T(\langle x_1, x_2\rangle) = \langle x_2, x_1\rangle$$

which reflects all vectors through the line $x_2 = x_1$. Does this linear transformation map any lines in R^2 to points or does it map all lines to lines? Explain.

Exercise 5.4.6. *In Section 5.3.5 we studied the shearing transformation*

$$T(\langle x_1, x_2\rangle) = \langle x_1 + x_2, x_2\rangle .$$

If you didn't quite picture the action of the shearing transformation when you studied Section 5.3.5, then perhaps this exercise will help you picture it better.

1. *Show that if L is a horizontal line in R^2, meaning that L has equation*

$$L : x_2 = k$$

where k is a constant, then $T(L) = L$.

2. *Suppose that L is the vertical line*

$$L : x_1 = 4.$$

Find $T(L)$. Then draw pictures of L and $T(L)$ in different colors.

3. *More generally, suppose that L is the vertical line*

$$L : x_1 = k$$

where k is a constant. Find $T(L)$.

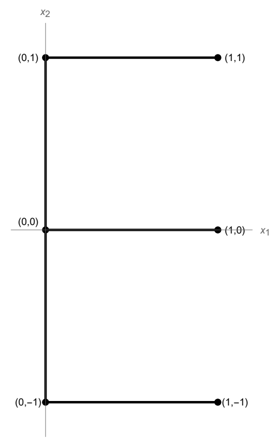


Figure 5.21: Letter E

Exercise 5.4.7. *Figure 5.21 shows a picture of capital letter "E". Shear the letter E using the shearing transformation* $T(\langle x_1, x_2\rangle) = \langle x_1 + x_2, x_2\rangle$ *and draw a picture of the sheared E.*

Having shown that linear transformations map lines to lines or points, we can actually say something more specific. If we have two lines, L_1 and L_2, in R^n that are *parallel* to each other, then any direction vector for L_1 is also a direction vector for L_2. If $\vec{d} \neq \vec{0}_n$ is a direction vector for L_1, then $\vec{d}$ is also a direction vector for L_2. This means that L_1 and L_2 have vector equations

$$L_1 : \vec{x} = \vec{p}_1 + t\vec{d}$$
$$L_2 : \vec{x} = \vec{p}_2 + t\vec{d}.$$

When we transform points on L_1 using a linear transformation T, we obtain

$$T(\vec{x}) = T(\vec{p}_1) + tT\left(\vec{d}\right)$$

and when we transform points on L_2 using this same linear transformation we obtain

$$T(\vec{x}) = T(\vec{p}_2) + tT\left(\vec{d}\right).$$

Thus, assuming that $T\left(\vec{d}\right) \neq \vec{0}_n$, we see that $T(L_1)$ has direction vector $T\left(\vec{d}\right)$ and $T(L_2)$ also has direction vector $T\left(\vec{d}\right)$, which means that the lines

$T(L_1)$ and $T(L_2)$ are parallel to each other. If it is the case that $T\left(\vec{d}\right) = \vec{0}_n$, then $T(L_1) = \{T(\vec{p}_1)\}$ and $T(L_2) = \{T(\vec{p}_2)\}$. This is summarized in the following corollary to Theorem 5.4.1.

Corollary 5.4.1. *Linear transformations map parallel lines to parallel lines (or to points). Specifically, if $T : R^n \to R^m$ is a linear transformation and L_1 and L_2 are the lines*

$$L_1 : \vec{x} = \vec{p}_1 + t\vec{d}$$
$$L_2 : \vec{x} = \vec{p}_2 + t\vec{d}$$

in R^n, where $\vec{d} \neq \vec{0}_n$, then

1. *If $\vec{d} \notin \ker(T)$, then $T(L_1)$ and $T(L_2)$ are parallel lines in R^m.*

2. *If $\vec{d} \in \ker(T)$, then $T(L_1)$ and $T(L_2)$ are points in R^m.*

Example 5.4.6. *In Example 5.4.3, we saw that the linear transformation $T(\langle x_1, x_2\rangle) = \langle -x_2, x_1\rangle$ rotates horizontal lines in R^2 to vertical lines in R^2. More generally, if L is a line with direction vector $\vec{d} = \langle d_1, d_2\rangle \neq \vec{0}_2$, then $T\left(\vec{d}\right) = \langle -d_2, d_1\rangle$ and we see that $T\left(\vec{d}\right)$ is orthogonal to $\vec{d}$. Hence T rotates any line, L, to a line that is perpendicular to L. This means that T maps parallel lines to parallel lines.*

In Example 5.4.5, we saw that the linear transformation $T(\langle x_1, x_2\rangle) = \langle x_1, 0\rangle$ projects all lines in R^2 onto the x_1 axis, with the exception of vertical lines. All non–vertical lines, whether parallel to each other or not, are projected onto the same line (the x_1 axis). Vertical lines are mapped to a single point on the x_1 axis. That is because a vertical line has direction vector $\vec{d} = \langle 0, 1\rangle$ and $T(\langle 0, 1\rangle) = \langle 0, 0\rangle$ so $\vec{d} \in \ker(T)$.

In Exercise 5.4.4, you were asked to show that the linear transformation $T(\vec{x}) = 2\vec{x}$ maps any line, L, to a line that is parallel to L. Thus T maps parallel lines to parallel lines.

5.5 Compositions and Similarity

The idea of composing two functions is applicable to functions in general and this idea is highly useful in studying linear transformations. In fact, we will see that our need to form compositions of linear transformations is the reason that we define matrix multiplication the way that we do.

Suppose that $f : D_f \to C_f$ is a function and suppose that $g : D_g \to C_g$ is some other function. Here we are denoting the domains of f and g by D_f and D_g respectively, and we are denoting the codomains of f and g by C_f and C_g respectively. If $\text{Range}(g)$ is a subset of D_f, then we can form a new function which is called a **composition**. This function, which is denoted by $f \circ g$ (and spoken as "f composition g") is the function that takes each element x in D_g, then maps it via g to $\text{Range}(g)$ to obtain the element $g(x)$, and then takes $g(x)$ and maps it via f to the element $f(g(x))$ in C_f. It is possible for us to form this function because we are assuming that $\text{Range}(g)$ is a subset of D_f, ensuring that for each $x \in D_g$ we have $g(x) \in D_f$. Thus $f \circ g$ is the function $f \circ g : D_g \to C_f$

$$(f \circ g)(x) = f(g(x)) \text{ for all } x \in D_g.$$

Figure 5.22 shows a schematic diagram of $f \circ g$.

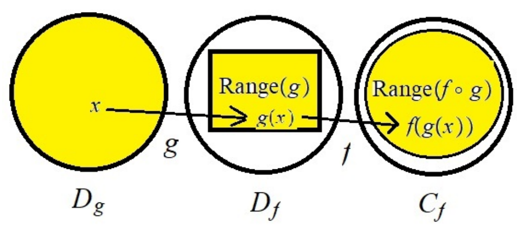


Figure 5.22: The Composition $f \circ g$

Function composition is something you have encountered in your previous math courses.[1] As an example, suppose that $g(x) = 2x$ and $f(x) = \sin(x)$. Then $f \circ g$ is the function

$$(f \circ g)(x) = f(g(x)) = \sin(g(x)) = \sin(2x)$$

and $g \circ f$ is the function

$$(g \circ f)(x) = g(f(x)) = 2f(x) = 2\sin(x).$$

[1]A topic that you probably remember from calculus where function composition comes into play is the Chain Rule, which tells us how to take the derivative of a composition of functions. It tells us that $(f \circ g)'(x) = f'(g(x))\, g'(x)$.

The above example shows that it is **not** true in general that $g \circ f = f \circ g$.

Forming and working with compositions of linear transformations $T : R^n \to R^m$ is particularly easy. To see why, let's start by just considering two linear functions $g : R \to R$ and $f : R \to R$ defined by $g(x) = bx$ and $f(x) = ax$ where a and b are some given constants. The composition $f \circ g$ is the function $f \circ g : R \to R$ defined by

$$(f \circ g)(x) = f(g(x)) = f(bx) = a(bx) = (ab)x$$

and the composition $g \circ f$ is the function $g \circ f : R \to R$ defined by

$$(g \circ f)(x) = g(f(x)) = g(ax) = b(ax) = (ba)x = (ab)x.$$

We see that in this case it is true that $g \circ f = f \circ g$.

Because of the way we have defined matrix multiplication, forming the composition of linear transformations from R^n to R^m works very similarly to forming compositions of linear functions from R to R. This is because the formula for any linear transformation from R^n to R^m can be written as the multiplication of a vector by a matrix. If $S : R^n \to R^p$ is the linear transformation defined by $S(\vec{x}) = A_S\vec{x}$ where A_S is a $p \times n$ matrix and $T : R^p \to R^m$ is the linear transformation defined by $T(\vec{x}) = A_T\vec{x}$ where A_T is an $m \times p$ matrix, then $T \circ S : R^n \to R^m$ is defined by

$$(T \circ S)(\vec{x}) = T(S(\vec{x})) = T(A_S\vec{x}) = A_T(A_S\vec{x}) = (A_TA_S)\vec{x}.$$

Hence if S has standard matrix A_S and T has standard matrix A_T, then $T \circ S$ has standard matrix A_TA_S. Notice that we have used the all–important associative property of matrix multiplication in deducing this fact. Also note that since matrix multiplication is not commutative, then even if $T \circ S$ and $S \circ T$ are both defined, it is **not** generally true that $T \circ S = S \circ T$.

Example 5.5.1. *Suppose that $S : R^3 \to R^2$ is the linear transformation*

$$S(\langle x_1, x_2, x_3\rangle) = \langle 2x_1 + x_2, 2x_1 + x_2 + x_3\rangle \tag{5.19}$$

and suppose that $T : R^2 \to R^3$ is the linear transformation

$$T(\langle x_1, x_2\rangle) = \langle -x_1, 3x_1 - x_2, -2x_1 + 3x_2\rangle\,. \tag{5.20}$$

1. *Fill in the blanks: $T \circ S$ is a linear transformation from ___ to ___.*

2. *Find the standard matrix for S and the standard matrix for T. Give the name A_S to the standard matrix for S and give the name A_T to the standard matrix for T.*

3. *Find the standard matrix for $T \circ S$.*

4. *Write the formula for $T \circ S$ in the form $(T \circ S)(\langle x_1, x_2, x_3 \rangle) = \_\_\_\_\_\_$.*

Solution:

1. *Since $S : R^3 \to R^2$ and Since $T : R^2 \to R^3$, then $T \circ S : R^3 \to R^3$.*

2. *Recall that the standard matrix for S is the matrix whose column vectors the images under S of the standard basis vectors of R^3. Since*

$$\begin{aligned} S(\langle 1,0,0\rangle) &= \langle 2(1)+0, 2(1)+0+0\rangle = \langle 2,2\rangle \\ S(\langle 0,1,0\rangle) &= \langle 2(0)+1, 2(0)+1+0\rangle = \langle 1,1\rangle \\ S(\langle 0,0,1\rangle) &= \langle 2(0)+0, 2(0)+0+1\rangle = \langle 0,1\rangle, \end{aligned}$$

then the standard matrix for S is

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

Likewise, the standard matrix for T is the matrix whose column vectors are the images under T of the standard basis vectors of R^2. Since

$$\begin{aligned} T(\langle 1,0\rangle) &= \langle -(1), 3(1)-0, -2(1)+3(0)\rangle = \langle -1,3,-2\rangle \\ T(\langle 0,1\rangle) &= \langle -0, 3(0)-1, -2(0)+3(1)\rangle = \langle 0,-1,3\rangle, \end{aligned}$$

then the standard matrix for T is

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

3. *The standard matrix for $T \circ S$ is*

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

4. *Since* $(T \circ S)(\vec{x}) = (A_T A_S)(\vec{x})$ *for all* $\vec{x} \in R^3$, *we have*

$$\begin{aligned}(T \circ S)(\vec{x}) &= (A_T A_S)(\vec{x}) \qquad (5.21)\\ &= \begin{bmatrix} -2 & -1 & 0 \\ 4 & 2 & -1 \\ 2 & 1 & 3 \end{bmatrix} \langle x_1, x_2, x_3 \rangle \\ &= \langle -2x_1 - x_2, 4x_1 + 2x_2 - x_3, 2x_1 + x_2 + 3x_3 \rangle .\end{aligned}$$

As a test to see if we have found the right formula for $T \circ S$, *let us take the randomly chosen vector* $\vec{x} = \langle -3, 2, 2 \rangle \in R^3$ *and compute* $(T \circ S)(\vec{x})$ *one step at a time using formulas (5.19) and (5.20) and then compute* $(T \circ S)(\vec{x})$ *in just one step by using formula (5.21). We should get the same answer either way.*

Using formulas (5.19) and (5.20), we obtain

$$S(\vec{x}) = S(\langle -3, 2, 2 \rangle) = \langle 2(-3) + 2, 2(-3) + 2 + 2 \rangle = \langle -4, -2 \rangle$$

and

$$\begin{aligned}T(S(\vec{x})) &= T(\langle -4, -2 \rangle) \\ &= \langle -(-4), 3(-4) - (-2), -2(-4) + 3(-2) \rangle \\ &= \langle 4, -10, 2 \rangle .\end{aligned}$$

Using (5.21) we obtain

$$\begin{aligned}(T \circ S)(\vec{x}) &= (T \circ S)(\langle -3, 2, 2 \rangle) \\ &= \langle -2(-3) - 2, 4(-3) + 2(2) - 2, 2(-3) + 2 + 3(2) \rangle \\ &= \langle 4, -10, 2 \rangle .\end{aligned}$$

Exercise 5.5.1. *For the linear transformations* S *and* T *given by formulas (5.19) and (5.20) in Example 5.5.1,*

1. *Fill in the blanks:* $S \circ T$ *is a linear transformation from* ___ *to* ___.

2. *The standard matrices for* S *and* T *have already been found in Example 5.5.1.*

3. *Find the standard matrix for* $S \circ T$.

4. *Write the formula for $S \circ T$ in the form $(S \circ T)(\langle x_1, x_2 \rangle) =$ _____.*

Exercise 5.5.2. *Let $S : R^2 \to R^2$ be the linear transformation defined by*

$$S(\langle x_1, x_2 \rangle) = \langle 0, x_2 \rangle .$$

This linear transformation projects all vectors in R^2 onto the x_2 axis.

Let $T : R^2 \to R^2$ be the linear transformation defined by

$$T(\langle x_1, x_2 \rangle) = \langle x_1, 0 \rangle .$$

This linear transformation projects all vectors in R^2 onto the x_1 axis.

1. *Without doing any calculations, what should $T \circ S$ do to all vectors in R^2? Explain in words.*

2. *Write down the matrix for S, the matrix for T, and the matrix for $T \circ S$.*

3. *Write down a formula for $T \circ S$ in the form $T\langle x_1, x_2 \rangle =$ ____. Does your answer agree with what you guessed in answering part 1?*

Exercise 5.5.3. *In Section 5.3.6, we studied the rotation transformations R_θ. Let $R_{45^\circ} : R^2 \to R^2$ be the linear transformation that rotates vectors in R^2 counterclockwise through angle 45°.*

1. *What does the linear transformation $R_{45^\circ} \circ R_{45^\circ}$ do to vectors in R^2? Explain in words and fill in the blank below*

$$R_{45^\circ} \circ R_{45^\circ} = \_\_\_\_\_ .$$

2. *Show that $A_{45^\circ} A_{45^\circ} = A_{90^\circ}$.*

3. *Without doing any computations, what do you guess that we get when we multiply the matrix A_{45° by itself eight times? In other words, what do you guess is $(A_{45^\circ})^8$. After guessing, compute $(A_{45^\circ})^8$ to see if your guess is correct.*

Exercise 5.5.4. *The linear transformation* $S : R^2 \to R^2$ *defined by*

$$S(\vec{x}) = 3\vec{x}$$

expands all vectors in R^2 *by a multiplier of* 3. *(This means that* $S(\vec{x})$ *points in the same direction as* $\vec{x}$ *and has* 3 *times the magnitude of* $\vec{x}$*).*

The linear transformation $T : R^2 \to R^2$ *defined by*

$$T(\vec{x}) = 4\vec{x}$$

expands all vectors in R^2 *by a multiplier of* 4. *(This means that* $T(\vec{x})$ *points in the same direction as* $\vec{x}$ *and has* 4 *times the magnitude of* $\vec{x}$*).*

1. *Explain in words what* $T \circ S$ *does to vectors in* R^2.
2. *Write down the matrices* A_S *and* A_T *and* $A_{T \circ S}$.
3. *Write a formula for* $T \circ S$ *in the form* $(T \circ S)(\vec{x}) =$ ______.
4. *Show that* $S \circ T = T \circ S$. *Does this make sense?*

Note that if $T : R^n \to R^n$ is an invertible linear transformation and T has matrix A_T, then T^{-1} has matrix $(A_T)^{-1}$. In other words, $A_{T^{-1}} = (A_T)^{-1}$. The **identity transformation** $E_n : R^n \to R^n$ is the transformation defined by

$$E_n(\vec{x}) = \vec{x} \text{ for all } \vec{x} \in R^n.$$

The identity transformation does *nothing* to all vectors in R^n. The matrix for E_n is I_n (the $n \times n$ identity matrix) because $E_n(\vec{x}) = \vec{x} = I_n\vec{x}$. Thus if T is any invertible linear transformation from R^n to R^n we have

$$\begin{aligned} T^{-1} \circ T &= E_n \\ T \circ T^{-1} &= E_n \\ A_{T^{-1}} A_T &= I_n \\ A_T A_{T^{-1}} &= I_n. \end{aligned}$$

Example 5.5.2. *Consider the linear transformation,* R_{45° *that rotates all vectors in* R^2 *counterclockwise by* 45°. *The matrix for this transformation is*

$$A_{45^\circ} = \begin{bmatrix} \cos(45^\circ) & -\sin(45^\circ) \\ \sin(45^\circ) & \cos(45^\circ) \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}.$$

How do we "undo" a rotation of 45° *counterclockwise? The answer is that we rotate* 45° *clockwise. In other words, we apply the linear transformation* R_{-45°*. The matrix for* R_{-45° *is*

$$A_{-45^\circ} = \begin{bmatrix} \cos(-45^\circ) & -\sin(-45^\circ) \\ \sin(-45^\circ) & \cos(-45^\circ) \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}.$$

Note that

$$A_{-45^\circ}A_{45^\circ} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} = I_2,$$

which is what we expect because R_{45° *and* R_{-45° *are inverses of each other.*

Exercise 5.5.5. *Let* R_{30° *be the linear transformation that rotates all vectors in* R^2 *by* 30° *clockwise. Write down the matrices for* R_{30° *and* $(R_{30^\circ})^{-1}$ *and verify that the product of these matrices (in either order) is* I_2*.*

Exercise 5.5.6. *We showed in Section 5.3.6 that the linear transformation that rotates all vectors in* R^2 *by angle* θ *is* $R_\theta(\vec{x}) = A_\theta \vec{x}$ *where* A_θ *is the matrix*

$$A_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

This tells us that the linear transformation $(R_\theta)^{-1} = R_{-\theta}$ *has matrix*

$$A_{-\theta} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Verify that this is correct by computing $A_{-\theta}A_\theta$*.*

The idea of function composition is very useful in studying linear transformations. By just being familiar with a few linear transformations like those we studied in Section 5.3, we can use the ones we are familiar with to construct new linear transformations by forming compositions. This is illustrated in the next two examples.

Example 5.5.3. *Let us construct the linear transformation* $T : R^2 \to R^2$ *that first rotates a vector in* R^2 *by* 60° *counterclockwise, then reflects the resulting vector through the* x_1 *axis, and then doubles the magnitude of the resulting vector.*

Solution: *We will construct T a piece at a time and then form the composition of the pieces.*

The first piece is the piece R_{60° that rotates vectors by 60° counterclockwise. This piece has matrix

$$A_{60^\circ} = \begin{bmatrix} \cos(60^\circ) & -\sin(60^\circ) \\ \sin(60^\circ) & \cos(60^\circ) \end{bmatrix} = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}.$$

The second piece is the reflection $S(\langle x_1, x_2\rangle) = \langle x_1, -x_2\rangle$ which has matrix

$$A_S = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The third piece is the expansion by a multiplier of 2 which is $P(\vec{x}) = 2\vec{x}$, which has matrix

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

The combined effect of first applying R_{60°, and then applying S, and then applying P is $T = P \circ S \circ R_{60^\circ}$, which has matrix

$$\begin{aligned} A_T &= A_P A_S A_{60^\circ} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & -1 \end{bmatrix}. \end{aligned}$$

Hence T is the linear transformation

$$T(\langle x_1, x_2\rangle) = A_T \langle x_1, x_2\rangle = \left\langle x_1 - \sqrt{3}x_2, -\sqrt{3}x_1 - x_2 \right\rangle.$$

Let us check some specific vectors and see if what we compute agrees with the visual picture we have of this transformation:

$$T(\langle 1, 0\rangle) = \left\langle 1 - \sqrt{3}(0), -\sqrt{3}(1) - 0 \right\rangle = \left\langle 1, -\sqrt{3} \right\rangle$$

seems to make sense because if we start with the vector $\langle 1, 0\rangle$ and then rotate it 60° counterclockwise, we end up in the first quadrant of R^2. If we then reflect *through the x_1 axis, we are in the fourth quadrant. Doubling the magnitude keeps us in the fourth quadrant. Note that $\left\langle 1, -\sqrt{3}\right\rangle$ is in the fourth quadrant.*

$$T\left(\langle 0,1\rangle\right) = \left\langle 0 - \sqrt{3}\,(1)\,, -\sqrt{3}\,(0) - 1\right\rangle = \left\langle -\sqrt{3}, -1\right\rangle$$

seems to make sense because if we start with the vector $\langle 0, 1\rangle$ and then rotate it $60°$ counterclockwise, we end up in the second quadrant. If we then reflect through the x_1 axis, we are in the third quadrant. Doubling the magnitude keeps us in the third quadrant. Note that $\left\langle -\sqrt{3}, -1\right\rangle$ is in the third quadrant.

Exercise 5.5.7. *In Example 5.5.3, we formed the composition $P \circ S \circ R_{60°}$. What if we had composed these in a different order? Would we get a different linear transformation than we got in Example 5.5.3. This exercise is aimed at investigating that question. Try to answer these questions without doing any calculations. Just try to visualize and maybe draw a few pictures to answer the questions. Then write down formulas for each of the transformations under investigation.*

1. *Is $P \circ R_{60°} \circ S$ the same or different from $P \circ S \circ R_{60°}$? In other words, is doing things in the order*

    - *reflect through x_1 axis*
    - *rotate by $60°$ counterclockwise*
    - *double magnitude*

    the same as doing things in the order

    - *rotate by $60°$ counterclockwise*
    - *reflect through x_1 axis*
    - *double magnitude?*

    Hint: Pick out a vector or two and visualize what quadrant that vector ends up in after each series of three successive transformations is applied.

2. *Is $S \circ R_{60°} \circ P$ the same or different from $P \circ S \circ R_{60°}$?*

3. *Is $S \circ P \circ R_{60°}$ the same or different from $P \circ S \circ R_{60°}$?*

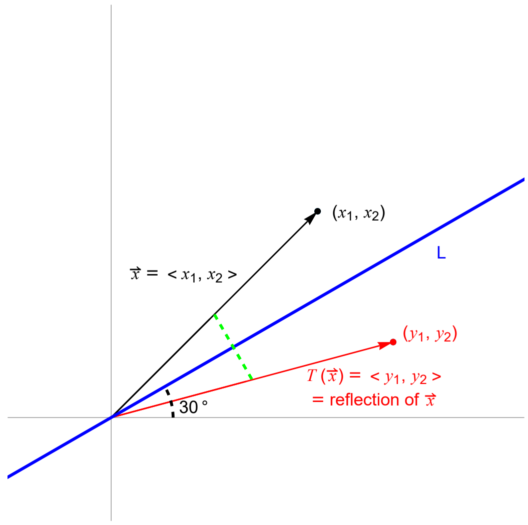


Figure 5.23: Reflecting Through L

Example 5.5.4. *Let us find the linear transformation* $T : R^2 \to R^2$ *that reflects all vectors in* R^2 *through the line,* L*, that makes an angle of* $30°$ *with the* x_1 *axis. See Figure 5.23.*

Solution: *Referring to Figure 5.23, observe that reflecting the vector* $\vec{x}$ *through the line* L *can be accomplished in three steps:*

a) First rotate the picture by $30°$ *clockwise.*

b) Then reflect through the x_1 *axis.*

c) Then rotate $30°$ *counterclockwise.*

The matrix for the linear transformation, $R_{-30°}$*, that rotates vectors* $30°$ *clockwise is*

$$A_{-30°} = \begin{bmatrix} \sqrt{3}/2 & 1/2 \\ -1/2 & \sqrt{3}/2 \end{bmatrix}.$$

The matrix for the linear transformation, S*, that reflects vectors through the* x_1 *axis, is*

$$A_S = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The matrix for the linear transformation, R_{30°, that rotates vectors 30° counterclockwise rotation is

$$A_{30^\circ} = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix}.$$

Thus the matrix for the linear transformation, $T = R_{30^\circ} \circ S \circ R_{-30^\circ}$ that reflects vectors through the line L is

$$A_T = A_{30^\circ} A_S A_{-30^\circ} = \begin{bmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{bmatrix}.$$

The formula for T is

$$T(\langle x_1, x_2 \rangle) = \left\langle \frac{1}{2}x_1 + \frac{\sqrt{3}}{2}x_2, \frac{\sqrt{3}}{2}x_1 - \frac{1}{2}x_2 \right\rangle.$$

The idea that was used in Example 5.5.4 was to relate a linear transformation that we were trying to find (reflection through the line, L, that makes an angle of 30° with the x_1 axis) to a "similar" linear transformation (reflection through the x_1 axis). Our strategy for doing this was

1. Map the problem to a simpler setting by rotating the whole picture 30° clockwise.

2. Solve the problem in the simpler setting. (The problem in the simpler setting is just to reflect through the x_1 axis.)

3. Map the problem back to the original setting by rotating the whole picture 30° counterclockwise.

More generally, suppose that we want to find some "complicated" linear transformation $T : R^2 \to R^2$. We could proceed as follows: Think of some "easier" linear transformation $S : R^2 \to R^2$ that is "similar" to T. (For example, maybe T is projection onto some arbitrary line, L, in R^2. Then we could choose S to be projection onto the x_1 axis.) Then

1. Map the problem to the simpler setting by applying some invertible linear transformation P.

2. Solve the problem in the simpler setting. (That is, find S.)

3. Then map the problem back to the original setting by applying P^{-1}.

The result is that we have

$$T = P^{-1} \circ S \circ P$$

where S is "simple" in comparison with T.

In the above discussion, we been have putting the words "simple" and "complicated" and "similar" in quotes, because we are using these terms informally. We have not given precise meanings to the terms. We can at least see, though, how we can give a precise meaning to the term "similar". We will do that now. The idea of similarity of linear transformations (and similarity of matrices) is important in the study of linear transformations.

Definition 5.5.1. *A linear transformation $T : R^n \to R^n$ is said to be* ***similar*** *to a linear transformation $S : R^n \to R^n$ if there exists an invertible linear transformation $P : R^n \to R^n$ such that $T = P^{-1} \circ S \circ P$.*

Likewise, an $n \times n$ matrix A is said to be ***similar*** *to an $n \times n$ matrix B, if there exists an invertible $n \times n$ matrix C such that $A = C^{-1}BC$.*

Note that if the linear transformation $T : R^n \to R^n$ is similar to the linear transformation $S : R^n \to R^n$, and A_T is the matrix for T and A_S is the matrix for S, then A_T is similar to A_S.

Example 5.5.5. *Let us show that the linear transformation $T : R^2 \to R^2$ that projects all vectors in R^2 onto the line $L : x_2 = x_1$ is similar to the linear transformation $S : R^2 \to R^2$ that projects all vectors in R^2 onto the x_1 axis.*

First note that S has a simple formula. It is $S(\langle x_1, x_2\rangle) = \langle x_1, 0\rangle$. Since the line $L : x_2 = x_1$ makes an angle of 45° with the positive x_1 axis, we can find T by first rotating the picture by 45° clockwise, then applying S, and then rotating the picture back to its original setting.

The matrix for 45° rotation clockwise is

$$R_{-45^\circ} = \begin{bmatrix} \cos(-45^\circ) & -\sin(-45^\circ) \\ \sin(-45^\circ) & \cos(-45^\circ) \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}.$$

The matrix for S is

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

The matrix for 45° rotation counterclockwise is

$$R_{45^\circ} = \begin{bmatrix} \cos(45^\circ) & -\sin(45^\circ) \\ \sin(45^\circ) & \cos(45^\circ) \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}.$$

Thus $T = R_{45^\circ} \circ S \circ R_{-45^\circ}$ *and the matrix for* T *is*

$$\begin{aligned} A_T &= A_{45^\circ} A_S A_{-45^\circ} \\ &= \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}. \end{aligned}$$

The formula for T *is*

$$T(\langle x_1, x_2 \rangle) = \left\langle \frac{1}{2}x_1 + \frac{1}{2}x_2, \frac{1}{2}x_1 + \frac{1}{2}x_2 \right\rangle.$$

We have shown that the linear transformation T *is similar to the linear transformation* S *and we have also shown that the matrix*

$$A_T = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \text{ is similar to the matrix } A_S = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

T *and* S *are both projection transformations, but they project onto different lines.* S *projects onto the* x_1 *axis and* T *projects onto the line* $L : x_2 = x_1$.

Observe that the matrix

$$A_S = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{5.22}$$

from Example 5.5.4 is simpler than the matrix

$$A_T = \begin{bmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{bmatrix} \tag{5.23}$$

from that example. Likewise, the matrix

$$A_S = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{5.24}$$

from Example 5.5.5 is simpler than the matrix

$$A_T = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{5.25}$$

from that example.

Why might we view the A_S matrices (5.22) and (5.24) as being "simpler" than the corresponding matrices A_T matrices (5.23) and (5.25)? Perhaps because they have "lots of zeros" and thus we get the intuitive sense that they are easier to work with than matrices that don't have "lots of zeros". In fact, the matrices in (5.22) and (5.24) are examples of what we refer to as diagonal matrices and they are, in many ways, easier to work with than matrices that are not diagonal matrices. In particular, it is easy to raise a diagonal matrix to a power (meaning to multiply the matrix by itself a certain number of times). To see an example of this, let us use the pair of matrices A_S given in (5.22) and A_T given in (5.23). We will compute A_S^2 and A_T^2.

The computation of A_S^2 is

$$A_S^2 = A_S A_S = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = I_2. \tag{5.26}$$

The fact that $A_S^2 = I_2$ makes sense because the composition $S \circ S$ is reflection through the x_1 axis followed by another reflection through the x_1 axis, which is the same as doing nothing. Hence $S \circ S$ is the identity transformation.

The computation of A_T^2 is

$$A_T^2 = A_T A_T = \begin{bmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{bmatrix} \begin{bmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{bmatrix} = I_2 \tag{5.27}$$

and this also makes sense because the composition $T \circ T$ is reflection through the line L, that makes an angle of $30°$ with the positive x_1 axis, followed by another reflection through the line L, which is the same as doing nothing. Hence $T \circ T$ is the identity transformation.

The matrix multiplication done in (5.26) requires less computation than the matrix multiplication done in (5.27) because two of the entries of A_S are 0 (and it is easy to multiply by 0) whereas A_T does not have any entries of 0. Let us formally define what we mean by a diagonal matrix and then state a theorem that tells us that it is easy to compute powers of a diagonal matrix.

Definition 5.5.2. *An $n \times n$ matrix A is said to be a* ***diagonal matrix*** *if the only non–zero entries of A are on its main diagonal. Specifically, $A = [a_{ij}]$ is a diagonal matrix if $a_{ij} = 0$ for all $i \neq j$.*

Example 5.5.6. *The matrix*

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

is a diagonal matrix and the matrix

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

is also a diagonal matrix.

Theorem 5.5.1. *If*

$$A = \begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

is a diagonal matrix, then for any integer $p \geq 1$ we have

$$A^p = \begin{bmatrix} a_{11}^p & 0 & 0 & \cdots & 0 \\ 0 & a_{22}^p & 0 & \cdots & 0 \\ 0 & 0 & a_{33}^p & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn}^p \end{bmatrix}.$$

Theorem 5.5.1 tells us that any power of a diagonal matrix, A, is also a diagonal matrix and the entries on the main diagonal of A^p are simply the entries on the main diagonal of A raised to the p power. For example if

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

then

$$A^2 = \begin{bmatrix} 3^2 & 0 \\ 0 & (-5)^2 \end{bmatrix} = \begin{bmatrix} 9 & 0 \\ 0 & 25 \end{bmatrix}$$

and

$$A^3 = \begin{bmatrix} 3^3 & 0 \\ 0 & (-5)^3 \end{bmatrix} = \begin{bmatrix} 27 & 0 \\ 0 & -125 \end{bmatrix}.$$

We will look at the problem of "diagonalizing" a given matrix in more detail in Chapter 6. This is the problem of deciding whether a given matrix A is similar to some diagonal matrix B or not. (We will see that it turns out that some matrices are diagonalizable and some are not.) We conclude this section with some exercises on the concept of similarity.

Exercise 5.5.8. *If we are given any $n \times n$ matrix B then it is easy to find another $n \times n$ matrix A that is similar to B. All we need to do is take any invertible $n \times n$ matrix C and let $A = C^{-1}BC$.*

1. *Let B be the matrix*
$$B = \begin{bmatrix} 1 & 1 \\ -3 & -3 \end{bmatrix}.$$
Find a matrix, A, that is similar to B.

2. *Let B be the matrix*
$$B = \begin{bmatrix} -2 & 3 & -1 \\ -1 & -3 & 2 \\ 3 & -1 & -2 \end{bmatrix}.$$
Find a matrix, A, that is similar to B.

Exercise 5.5.9.

1. *Explain why any $n \times n$ matrix, A, is similar to itself.*
2. *Explain why if A is similar to B, then B is similar to A.*
3. *Explain why if A is similar to B and B is similar to C, then A is similar to C.*

Exercise 5.5.10. *Explain why the only matrix that is similar to*

$$O_{2\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

is $O_{2\times 2}$.

Exercise 5.5.11. *Explain why the only linear transformation from R^2 to R^2 that is similar to the identity transformation $E(\vec{x}) = \vec{x}$ is E itself.*

Exercise 5.5.12. *Explain why the only linear transformation from R^2 to R^2 that is similar to the linear transformation $T(\vec{x}) = 2\vec{x}$ is T itself.*

5.6 Linear Transformations for General Vector Spaces

Thus far we have considered only linear transformations $T : R^n \to R^m$. Since all such transformations have the form $T(\vec{x}) = A\vec{x}$ where A is some $m \times n$ matrix, we have the tools that we need (matrix algebra) to be able to understand these linear transformations very well. We can answer questions regarding the range and kernel of T by using A and can also determine whether or not T is invertible. We can easily form the composition of two linear transformations by multiplying their matrices. We will now be more general and define what we mean by a linear transformation $T : V \to W$ where V and W can be any vector spaces. It will be seen that all of the central concepts that apply to linear transformations $T : R^n \to R^m$ such as range, kernel, and invertibility, carry over in a natural way to more general linear transformations $T : V \to W$. The only difference is that it does not any longer make sense to write $T(\vec{x}) = A\vec{x}$, where A is a matrix, because the vectors in V are not assumed to be vectors in R^n. We will see, however, that when the vector spaces V and W are finite dimensional, we can still use matrix tools to study $T : V \to W$ by working with coordinate vectors.

The definition of a linear transformation $T : V \to W$ parallels Definition 5.2.1.

Definition 5.6.1. *Suppose that V and W are vector spaces. A linear transformation from V to W is a function $T : V \to W$ that has the properties:*

1. *If $\vec{x}$ and $\vec{y}$ are any two vectors in V, then $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$.*

2. *If $\vec{x}$ is any vector in V and c is any scalar, then $T(c\vec{x}) = cT(\vec{x})$.*

We have already seen many examples of linear transformations $T : R^n \to R^m$. Let us now look at two examples that involve other vector spaces.

Example 5.6.1. *Let $S : R^\infty \to R^\infty$ be the function defined by*

$$S\left(\langle a_1, a_2, a_3, \ldots\rangle\right) = \langle a_2, a_3, a_4, \ldots\rangle .$$

This is called a shifting transformation because the action of S is to discard the first component of the input and then shift all other components one position to the left. We will use Definition 5.6.1 to prove that S is a linear transformation. We will only prove that the first requirement of Definition 5.6.1 is satisfied and leave it as an exercise for you to prove that the second requirement of the definition is also satisfied.

Suppose that $\vec{a} = \langle a_1, a_2, a_3, \ldots\rangle$ and $\vec{b} = \langle b_1, b_2, b_3, \ldots\rangle$ are vectors in R^∞. Then

$$\begin{aligned} S\left(\vec{a} + \vec{b}\right) &= S\left(\langle a_1, a_2, a_3, \ldots\rangle + \langle b_1, b_2, b_3, \ldots\rangle\right) \\ &= S\left(\langle a_1 + b_1, a_2 + b_2, a_3 + b_3, \ldots\rangle\right) \\ &= \langle a_2 + b_2, a_3 + b_3, a_4 + b_4, \ldots\rangle \\ &= \langle a_2, a_3, a_4, \ldots\rangle + \langle b_2, b_3, b_4, \ldots\rangle \\ &= S\left(\vec{a}\right) + S\left(\vec{b}\right). \end{aligned}$$

This shows that S satisfies the first requirement of Definition 5.6.1.

Exercise 5.6.1. *Show that the shifting transformation, S, defined in Example 5.6.1 satisfies the second requirement of Definition 5.6.1.*

Example 5.6.2. *In Calculus I, you learned about derivatives. The process of taking the derivative of a function is called differentiation. You probably did not realize at the time that differentiation is a linear transformation! The reason (which you learned in Calculus I) is that*

1. *If f and g are two differentiable functions, then $f + g$ is also a differentiable function and*
$$(f + g)' = f' + g'$$
and

2. *If f is a differentiable function and c is a constant (scalar), then cf is also a differentiable function and*
$$(cf)' = cf'.$$

To formalize this fact, we define $D : C^1(R) \to C^0(R)$ to be the function

$$D(f) = f'.$$

Then D is a linear transformation because for all f and g in $C^1(R)$ and all scalars c we have both

$$D(f+g) = D(f) + D(g)$$

and

$$D(cf) = cD(f),$$

showing that both requirements of Definition 5.6.1 are satisfied.

To illustrate with a specific example, suppose that f is the function defined by $f(x) = x^3$ and g is the function defined by $g(x) = \sin(x)$. Then

$$\begin{aligned} D(f+g) &= D\left(x^3 + \sin(x)\right) \\ &= 3x^2 + \cos(x) \\ &= D(f) + D(g) \end{aligned}$$

and

$$\begin{aligned} D(3f) &= D\left(3x^3\right) \\ &= 9x^2 \\ &= 3\left(3x^2\right) \\ &= 3D(f). \end{aligned}$$

Exercise 5.6.2. *For the differentiation transformation $D(f) = f'$, compute*

1. $D(\cos(x))$
2. $D(x^4 - 1)$
3. $D(5x^3 - 2x + 16)$
4. $D(e^x)$
5. $D\left(-\frac{3}{2}e^x \sin(x) - \frac{7}{2}e^x \cos(x)\right)$

Exercise 5.6.3. *Let $M_{2\times 2}$ be the vector space of all 2×2 matrices with real entries. Let $f : M_{2\times 2} \to M_{2\times 2}$ be the function defined by $f(A) = \operatorname{rref}(A)$. Is f a linear transformation? Explain why or why not.*

5.6.1 Range, Kernel, and Invertibility

The definitions of the range and the kernel of a linear transformation $T : V \to W$ are easily generalized from the definitions that we gave for these concepts for linear transformations $T : R^n \to R^m$.

Definition 5.6.2. *If V and W are vector spaces and $T : V \to W$ is a linear transformation, we define the* ***range*** *of T to be*

$$\text{Range}(T) = \{\vec{y} \in W \mid T(\vec{x}) = \vec{y} \textit{ for at least one } \vec{x} \in V\}$$

and we define the ***kernel*** *(also called* ***null space****) of T to be*

$$\ker(T) = \left\{\vec{x} \in V \mid T(\vec{x}) = \vec{0}_W\right\}.$$

Note that two alternative ways to describe Range (T) are

$$\text{Range}(T) = \{T(\vec{x}) \mid \vec{x} \in V\}$$

and

$$\text{Range}(T) = T(V).$$

Just as was the case for linear transformations $T : V \to W$, it is always true that Range (T) is a subspace of W and that $\ker(T)$ is a subspace of V. You are asked to prove these facts in Exercises 5.6.5 and 5.6.6. Furthermore, it is also true that if V is a finite dimensional vector space, then the analog of Theorem 5.2.3 (Fundamental Theorem of Linear Algebra) holds. Specifically if V is finite dimensional, then

$$\dim(\text{Range}(T)) + \dim(\ker(T)) = \dim(V).$$

The concept of invertibility of a linear transformation $T : R^n \to R^m$ also carries over to linear transformations $T : V \to W$. In fact this concept was defined more generally for functions (not necessarily linear transformations) $f : A \to B$ in Section 5.1.2.

Definition 5.6.3. *Suppose that V and W are vector spaces and suppose that $T : V \to W$ is a linear transformation. We say that T is* ***invertible*** *if* Range $(T) = W$ *and T is also one–to–one. If T is invertible, then the* ***inverse*** *of T is defined to be the function $T^{-1} : W \to V$ defined by*

$$T^{-1}(\vec{y}) = \vec{x} \textit{ where } \vec{x} \textit{ is the unique vector in } V \textit{ such that } T(\vec{x}) = \vec{y}.$$

Note that if $T : V \to W$ is invertible, then

$$\left(T^{-1} \circ T\right)(\vec{x}) = T^{-1}\left(T\left(\vec{x}\right)\right) = \vec{x} \text{ for all } \vec{x} \in V$$

and

$$\left(T \circ T^{-1}\right)(\vec{x}) = T\left(T^{-1}\left(\vec{x}\right)\right) = \vec{x} \text{ for all } \vec{x} \in W.$$

For an invertible linear transformation $T : R^n \to R^m$, it was easily seen that T^{-1} is also a linear transformation. That was because we saw that if A is the matrix of T, then A^{-1} is the matrix of T^{-1}. When dealing with linear transformations in general, we don't have a formula of the form $T(\vec{x}) = A\vec{x}$ to work with. Nonetheless, it is always true that the inverse of an invertible linear transformation is a linear transformation. This is stated in the following theorem.

Theorem 5.6.1. *Suppose that V and W are vector spaces and suppose that $T : V \to W$ is an invertible linear transformation. Then $T^{-1} : W \to V$ as defined in Definition 5.6.3 is also a linear transformation.*

Proof. We need to show that T^{-1} satisfies both of the requirements of Definition 5.6.1. Suppose that $\vec{x}$ and $\vec{y}$ are vectors in W. Since T is invertible, there is a unique vector $\vec{u}$ in V such that $T(\vec{u}) = \vec{x}$ and hence $\vec{u} = T^{-1}(\vec{x})$. Likewise, there is a unique vector $\vec{v}$ in V such that $T(\vec{v}) = \vec{y}$ and hence $\vec{v} = T^{-1}(\vec{y})$. Since T is a linear transformation then

$$T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v}) = \vec{x} + \vec{y}$$

which implies that

$$T^{-1}(\vec{x} + \vec{y}) = \vec{u} + \vec{v} = T^{-1}(\vec{x}) + T^{-1}(\vec{y}).$$

This shows that T^{-1} satisfies the first requirement of Definition 5.6.1. In Exercise 5.6.4, you are asked to show that T^{-1} satisfies the second requirement of Definition 5.6.1. □

Exercise 5.6.4. *Complete the proof of Theorem 5.6.1 by showing that T^{-1} satisfies the second requirement of Definition 5.6.1.*

Exercise 5.6.5. *Suppose that $T : V \to W$ is a linear transformation. Prove that* Range (T) *is a subspace of W.*

Exercise 5.6.6. *Suppose that $T : V \to W$ is a linear transformation. Prove that* $\ker(T)$ *is a subspace of* V*.*

Example 5.6.3. *In Example 5.6.1, we considered the shifting transformation* $S : R^\infty \to R^\infty$ *defined by*

$$S(\langle a_1, a_2, a_3, \ldots \rangle) = \langle a_2, a_3, a_4, \ldots \rangle .$$

What is the range of S? In other words, what is the set of all vectors $\vec{b} \in R^\infty$ for which there exists a vector $\vec{a} \in R^\infty$ such that $S(\vec{a}) = \vec{b}$? After a moment of thought, we can see that $\text{Range}(S) = R^\infty$*. If we take any vector $\vec{b} = \langle b_1, b_2, b_3, \ldots \rangle \in R^\infty$, then $\vec{b}$ is the image under S of the vector $\vec{a} = \langle 0, b_1, b_2, b_3, \ldots \rangle$ because*

$$S(\vec{a}) = S(\langle 0, b_1, b_2, b_3, \ldots \rangle) = \langle b_1, b_2, b_3, \ldots \rangle = \vec{b}.$$

What is the kernel of S? In other words, what is the set of all vectors $\vec{a} \in R^\infty$ for which $S(\vec{a}) = \vec{0}$? (Recall that in R^∞, the zero vector is $\vec{0} = \langle 0, 0, 0, \ldots \rangle$.) After some thought, we see that $\ker(S)$ *consists of all vectors $\vec{a} \in R^\infty$ that have the form*

$$\vec{a} = \langle a_1, 0, 0, 0, \ldots \rangle$$

(meaning that all components of $\vec{a}$ are 0 except possibly for the first component). For such vectors $\vec{a}$ we see that

$$S(\vec{a}) = S(\langle a_1, 0, 0, 0, \ldots \rangle) = \langle 0, 0, 0, \ldots \rangle = \vec{0}.$$

Furthermore, if the vector $\vec{a}$ has a nonzero component anywhere beyond the first component, then $S(\vec{a}) \neq \vec{0}$. We can write $\ker(S)$ *as*

$$\ker(S) = \text{Span}\{\langle 1, 0, 0, 0, \ldots \rangle\} .$$

Exercise 5.6.7. *In Example 5.6.3, we discovered that the range and kernel of the shifting transformation, S, are*

$$\begin{aligned} \text{Range}(S) &= R^\infty \\ \ker(S) &= \text{Span}\{\langle 1, 0, 0, 0, \ldots \rangle\} . \end{aligned}$$

The fact that $\text{Range}(S) = R^\infty$ *means that S maps R^∞ onto R^∞. Explain why S is not one–to–one and hence not invertible.*

Example 5.6.4. *In Example 5.6.2 we studied the differentiation transformation $D : C^1(R) \to C^0(R)$ defined by*

$$D(f) = f'.$$

*Recall that $C^1(R)$ is the set of all functions with domain R that are differentiable and have **continuous** derivatives. That is why it was okay for us to designate the codomain of D to be $C^0(R)$, which is the set of all continuous functions that have domain R. If f has a continuous derivative, then $D(f) = f'$ is a continuous function.*[2]

You may recall that one part of the Fundamental Theorem of Calculus (studied in Calculus I) tells us that any continuous function has an antiderivative. This means that if g is any continuous function, then there exists a differentiable function f such that $D(f) = g$. Therefore the range of D is

$$\mathrm{Range}(D) = C^0(R).$$

Exercise 5.6.8. *In Example 5.6.4 we explained why the differentiation transformation, $D(f) = f'$ has range* Range $(D) = C^0(R)$. *Explain why D is not one–to–one and hence not invertible.*

Hint: How many solutions does the equation $D(f) = 5$ have?

Example 5.6.5. *In this example we look at the differentiation transformation, but with domain restricted to $P_2 = \mathrm{Span}\{1, x, x^2\}$. The elements of P_2 are polynomial functions that have degree 2 or less. Thus the elements of P_2 have the form*

$$p(x) = a_0 + a_1 x + a_2 x^2 \tag{5.28}$$

where a_0, a_1, and a_2 can be any scalars. When we take the derivative of such a function we obtain a polynomial function of degree 1 or less:

$$D(p) = a_1 + 2a_2 x. \tag{5.29}$$

*This means that if want to study the transformation $D(p) = p'$ with domain P_2, we can choose the codomain to be P_1. We might as well do that, since then we have that D maps P_2 **onto** P_1. In other words* Range $(D) = P_1$. *To be sure we see why* Range $(D) = P_1$, *note that if*

$$q(x) = b_0 + b_1 x \in P_1,$$

[2] There do exist functions that are differentiable but whose derivatives are not continuous. These functions are studied in other courses.

then $q = D(p)$ *where* $p \in P_2$ *is the function*

$$p(x) = b_0 x + \frac{1}{2} b_1 x^2.$$

Is D *one–to–one? The answer is no. To see why not, let* $q \in P_1$ *be the function*

$$q(x) = 5 + 4x.$$

There are infinitely many functions $p \in P_2$ *such that* $D(p) = q$*. For example*

$$D\left(5x + 2x^2\right) = 5 + 4x$$

and

$$D\left(9 + 5x + 2x^2\right) = 5 + 4x.$$

We conclude that D *is not invertible, because* D *is not one–to–one.*

Examining this from the point of view of Fundamental Theorem of Linear Algebra, since $\dim(\text{Range}(D)) = \dim(P_1) = 2$*, then we must have*

$$2 + \dim(\ker(D)) = \dim(P_2) = 3$$

and hence we must have $\dim(\ker(D)) = 1$*. To see why* $\dim(\ker(D)) = 1$*, note that* $\ker(D)$ *consists of all functions* $p \in P_2$ *such that* $D(p) = z$ *where* z *is the zero polynomial (meaning the polynomial such that* $z(x) = 0$ *for all* $x \in R$*). The only functions in* P_2 *that have derivative* z *are the constant functions. Thus* $\ker(D) = \text{Span}\{1\}$*, meaning that* $\dim(\ker(D)) = 1$*.*

Exercise 5.6.9. *In Example 5.6.5 we saw that the differentiation transformation* $D : P_2 \to P_1$ *defined by* $D(p) = p'$ *maps* P_2 *onto* P_1 *but is not one–to–one and hence not invertible. Let* $P_2^* = \text{Span}\{x, x^2\}$*. Thus* P_2^* *consists of all polynomial functions of the form*

$$p(x) = a_1 x + a_2 x^2.$$

Since the derivative of such a function has the form

$$D(p) = a_1 + 2a_2 x,$$

then we can choose our codomain to be P_1*.*

Let $D : P_2^* \to P_1$ *be the differentiation transformation* $D(p) = p'$*.*

1. *Explain why D is one–to–one and hence invertible.*

2. *Determine* $\text{Range}(D)$ *and* $\ker(D)$ *and show that the Fundamental Theorem of Linear Algebra*

$$\dim(\text{Range}(D)) + \dim(\ker(D)) = \dim(P_2^*)$$

is satisfied.

5.6.2 Powers of Linear Transformations $T : V \to V$

From this point on we will consider only linear transformations $T : V \to V$, where the domain and codomain of T are the same vector space. For such linear transformations, we have $\text{Range}(T) \subseteq V$ and thus we can compute $T \circ T$. In fact, since $\text{Range}(T \circ T) \subseteq V$, we can compute $T \circ T \circ T$ and so on. We can compose T with itself as many times as we like. The standard notation that is used when composing a linear transformation with itself n times is T^n. Thus

$$\begin{aligned} T^2 &= T \circ T \\ T^3 &= T \circ T \circ T \\ &\vdots \\ &\text{etc.} \end{aligned}$$

Exercise 5.6.10. *Let $S : R^\infty \to R^\infty$ be the shifting transformation*

$$S(\langle a_1, a_2, a_3, \ldots \rangle) = \langle a_2, a_3, a_4, \ldots \rangle$$

that was introduced in Example 5.6.1. Give formulas for S^2 and S^3.

$$\begin{aligned} S^2(\langle a_1, a_2, a_3, \ldots \rangle) &= \_\_\_\_\_\_\_\_\_\_ \\ S^3(\langle a_1, a_2, a_3, \ldots \rangle) &= \_\_\_\_\_\_\_\_\_\_. \end{aligned}$$

Hint: Remember that $S^2 = S \circ S$. This means that

$$S^2(\vec{a}) = (S \circ S)(\vec{a}) = S(S(\vec{a})).$$

Exercise 5.6.11. *Let $D : C^\infty(R) \to C^\infty(R)$ be the differentiation transformation $D(f) = f'$. This is the same transformation that was introduced* *in Example 5.6.2, except that we are now restricting the domain to include only functions that have derivatives of all orders; i.e., functions whose first derivative, second derivative, third derivative....and all derivatives exist. In restricting the domain of D to be $C^{\infty}(R)$, the linear transformation D now maps its domain into its domain and thus it is possible to form any number of compositions of D with itself. As an example, we have*

$$\begin{aligned} D\left(x^3\right) &= 3x^2 \\ D^2\left(x^3\right) &= D\left(D\left(x^3\right)\right) = 6x \\ D^3\left(x^3\right) &= D\left(6x\right) = 6 \\ D^4\left(x^3\right) &= D\left(6\right) = 0. \end{aligned}$$

We see that $D^n\left(x^3\right) = 0$ for all $n \geq 4$.

1. *Find $D^n\left(x^4 - 2x^3 + 5x - 2\right)$ for all $n = 1, 2, 3, \ldots$*

2. *Find $D^n\left(\sin\left(x\right)\right)$ for $n = 1, 2, 3$ and 4.*

3. *Find $D^n\left(e^x\right)$ for all $n = 1, 2, 3, \ldots$*

4. *Find $D^n\left(xe^x\right)$ for $n = 1, 2$ and 3.*

5. *Find $D^n\left(x\sin\left(x\right)\right)$ for $n = 1, 2, 3$ and 4.*

5.6.3 Working with Coordinate Vectors

We will now consider linear transformations $T : S \to S$ where S is a finite dimensional subspace of a vector space V. The vector space V itself can be either finite dimensional or infinite dimensional but we will only be using a finite dimensional subspace of V as the domain of our linear transformation and we will only consider linear transformations that map this finite dimensional subspace into itself. Hence we will be able to discuss composition of T with itself any number of times as in Section 5.6.2.

If S is a finite dimensional subspace of the vector space V and $\mathcal{B} = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is an ordered basis for S consisting of k vectors (where $k \geq 1$), then the coordinate vector of any vector $\vec{x} \in S$ with respect to the ordered basis $\mathcal{B}$ is the vector in R^k whose components are the unique weights that are used in writing $\vec{x}$ as a linear combination of the vectors in $\mathcal{B}$. In other words,

$$\vec{x} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k \qquad (5.30)$$

if and only if

$$[\vec{x}]_{\mathcal{B}} = \langle c_1, c_2, \ldots, c_k \rangle \,. \tag{5.31}$$

Lemma 4.8.1 in Section 4.8 tells us that if $\vec{x}$ and $\vec{y}$ are any two vectors in S and c is any scalar, then

$$[\vec{x} + \vec{y}]_{\mathcal{B}} = [\vec{x}]_{\mathcal{B}} + [\vec{y}]_{\mathcal{B}} \tag{5.32}$$

and

$$[c\vec{x}]_{\mathcal{B}} = c\,[\vec{x}]_{\mathcal{B}} \,. \tag{5.33}$$

The facts (5.32) and (5.33) tell us that the mapping from S into R^k that maps each vector $\vec{x} \in S$ to its coordinate vector $[\vec{x}]_{\mathcal{B}}$ is a linear transformation! Furthermore, this linear transformation is invertible because (5.30) is true if and only if (5.31) is true. We will refer to the linear transformation that maps a vector $\vec{x} \in S$ to its coordinate vector with respect to basis $\mathcal{B}$ as a **coordinate mapping**. Rather than giving a capital letter name to the coordinate mapping (as we normally do for linear transformations), we will just use the bracket symbol $[\cdot]_{\mathcal{B}}$ to denote the coordinate mapping and we will use the symbol $[\cdot]_{\mathcal{B}}^{-1}$ to denoted the inverse of the coordinate mapping. Thus $[\cdot]_{\mathcal{B}} : S \to R^k$ is the linear transformation defined by

$$[\vec{x}]_{\mathcal{B}} = \langle c_1, c_2, \ldots, c_k \rangle$$

where $\langle c_1, c_2, \ldots, c_k \rangle$ is the unique vector in R^k that satisfies (5.30), and $[\cdot]_{\mathcal{B}}^{-1} : R^k \to S$ is the linear transformation defined by

$$[\langle c_1, c_2, \ldots, c_k \rangle]_{\mathcal{B}}^{-1} = \vec{x}$$

where $\vec{x}$ is the unique vector in S that satisfies (5.30). The dot $(\cdot)$ that appears in the notations $[\cdot]_{\mathcal{B}}$ and $[\cdot]_{\mathcal{B}}^{-1}$ is just a place holder. If we plug some specific vector into $[\cdot]_{\mathcal{B}}$ or $[\cdot]_{\mathcal{B}}^{-1}$, then that vector takes the place of the dot. This kind of notation is commonly used in mathematics in situations where such a notation is convenient. Before proceeding, let us look at some examples to make sure that we understand the bracket notation for the coordinate mapping and inverse coordinate mapping (because we will be using this notation a lot in what follows).

Example 5.6.6. *We showed in Example 4.7.7 that the set of vectors* $\mathcal{B} = \{1, x, x^2\}$ *is linearly independent and is thus an ordered basis for the vector* *space* $P_2 = \mathrm{Span}\{1, x, x^2\}$. *The elements of* P_2 *are polynomial functions of the form*

$$p(x) = a_0 + a_1 x + a_2 x^2.$$

The coordinate vector of the above function p *with respect to the ordered basis* $\mathcal{B}$ *is*

$$[p(x)]_{\mathcal{B}} = \langle a_0, a_1, a_2 \rangle .$$

If $\langle c_0, c_1, c_2 \rangle$ *is any vector in* R^3, *then the inverse coordinate vector of* $\langle c_0, c_1, c_2 \rangle$ *is the function*

$$[\langle c_0, c_1, c_2 \rangle]^{-1} = c_0 + c_1 x + c_2 x^2.$$

As specific examples,

$$\left[-5 + 2x + 3x^2\right]_{\mathcal{B}} = \langle -5, 2, 3 \rangle$$

and

$$[\langle -4, -1, -5 \rangle]_{\mathcal{B}}^{-1} = -4 - x - 5x^2.$$

Example 5.6.7. *Consider the vector space* R^∞. *We can obtain a two dimensional subspace of* R^∞ *by choosing any two vectors,* $\vec{a}$ *and* $\vec{b}$, *in* R^∞ *such that* $\mathcal{B} = \left\{\vec{a}, \vec{b}\right\}$ *is a linearly independent set and then taking our subspace to be* $S = \mathrm{Span}\left\{\vec{a}, \vec{b}\right\}$. *Let us choose*

$$\begin{aligned} \vec{a} &= \langle 1, 0, 1, 0, 1, 0, \ldots \rangle \text{ (alternating 1 and 0 starting with 1)} \\ \vec{b} &= \langle 0, 1, 0, 1, 0, 1, \ldots \rangle \text{ (alternating 1 and 0 starting with 0).} \end{aligned}$$

The set $\mathcal{B} = \left\{\vec{a}, \vec{b}\right\}$ *is linearly independent because it is a set of two vectors and neither one of them is a scalar multiple of the other one. Thus* $\mathcal{B} = \left\{\vec{a}, \vec{b}\right\}$ *is an ordered basis for the two dimensional subspace* $S = \mathrm{Span}\left\{\vec{a}, \vec{b}\right\}$.

Let $\vec{x}$ *be the vector*

$$\vec{x} = \langle -2, 2, -2, 2, -2, 2, \ldots \rangle .$$

Then $\vec{x} \in S$ *because* $\vec{x} = -2\vec{a} + 2\vec{b}$. *The coordinate vector of* $\vec{x}$ *with respect to the ordered basis* $\mathcal{B}$ *is*

$$[\vec{x}]_{\mathcal{B}} = \langle -2, 2 \rangle$$

and the inverse coordinate vector of $\langle -2, 2\rangle$ *is*

$$[\langle -2, 2\rangle]_{\mathcal{B}}^{-1} = \vec{x}.$$

Both of the above two facts are due to the fact that $\vec{x} = -2\vec{a} + 2\vec{b}$.

Example 5.6.8. *Let* $S = \operatorname{Span}\{\sin(x), \cos(x)\}$. *We showed in Example 4.7.12 that the set of functions* $\mathcal{B} = \{\sin(x), \cos(x)\}$ *is linearly independent. Thus* $\mathcal{B}$ *is a basis for* S. *The coordinate vector of the function* $-3\sin(x) + \cos(x)$ *with respect to the ordered basis is*

$$[-3\sin(x) + \cos(x)]_{\mathcal{B}} = \langle -3, 1\rangle$$

and the inverse coordinate vector of the vector $\langle -3, 1\rangle$ *is*

$$[\langle -3, 1\rangle]_{\mathcal{B}}^{-1} = -3\sin(x) + \cos(x).$$

Exercise 5.6.12. *For the vector space* $P_2 = \operatorname{Span}\{1, x, x^2\}$ *with ordered basis* $\mathcal{B} = \{1, x, x^2\}$, *find*

a) $[2 + 2x + 2x^2]_{\mathcal{B}}$

b) $[1]_{\mathcal{B}}$

c) $[x]_{\mathcal{B}}$

d) $[x^2]_{\mathcal{B}}$

e) $[\langle 1, -2, -2\rangle]_{\mathcal{B}}^{-1}$

f) $[\langle 0, 1, 1\rangle]_{\mathcal{B}}^{-1}$

Exercise 5.6.13. *Let* $\vec{a}$ *and* $\vec{b}$ *be the vectors given in Example 5.6.7. We showed in that example that* $\mathcal{B} = \left\{\vec{a}, \vec{b}\right\}$ *is linearly independent and is thus an ordered basis for* $S = \operatorname{Span}\left\{\vec{a}, \vec{b}\right\}$. *Find*

a) $[\langle 0, 0, 0, 0, 0, 0, 0, \ldots\rangle]_{\mathcal{B}}$

b) $[\langle 1, 2, 1, 2, 1, 2, 1, \ldots\rangle]_{\mathcal{B}}$

c) $[\langle 2, -4\rangle]_{\mathcal{B}}^{-1}$

Exercise 5.6.14. *The set of vectors*

$$\mathcal{B} = \{\sin(x), \cos(x), x\sin(x), x\cos(x)\}$$

is linearly independent and is thus a basis for

$$S = \operatorname{Span}\{\sin(x), \cos(x), x\sin(x), x\cos(x)\}.$$

Find

a) $[-3\sin(x) + 2\cos(x) + 3x\sin(x) - 3x\cos(x)]_{\mathcal{B}}$

b) $[\cos(x)]_{\mathcal{B}}$

c) $[-4\cos(x) + 5x\sin(x) + x\cos(x) + 3\sin(x)]_{\mathcal{B}}$

d) $[\langle -4, 2, 4, -1\rangle]_{\mathcal{B}}^{-1}$

1. $[\langle 0, 0, 1, 0\rangle]_{\mathcal{B}}^{-1}$

We will now see how we can use coordinate vectors as a handy tool for studying a linear transformation $T : S \to S$ where S is finite dimensional and $\mathcal{B} = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is an ordered basis for S. The idea, which is described precisely in Theorem 5.6.2, is to consider a similar linear transformation $T_{\mathcal{B}} : R^k \to R^k$ that mimics T. The advantage of working with $T_{\mathcal{B}}$ is that we have a matrix to work with because $T_{\mathcal{B}}$ has the form $T_{\mathcal{B}}(\vec{c}) = A_{\mathcal{B}}\vec{c}$ for some $k \times k$ matrix $A_{\mathcal{B}}$. Theorem 5.6.2 tells us how to find the matrix $A_{\mathcal{B}}$, which is called the **matrix of the linear transformation** T **with respect to the ordered basis** B.

Theorem 5.6.2. *Suppose that S is a finite dimensional subspace of a vector space V (where V may be finite dimensional or infinite dimensional) and suppose that $T : S \to S$ is a linear transformation. Suppose also that $\mathcal{B} = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is an ordered basis of S. Define $A_{\mathcal{B}}$ to be the $k \times k$ matrix whose columns are*

$$\operatorname{Col}_j(A_{\mathcal{B}}) = [T(\vec{u}_j)]_{\mathcal{B}} \tag{5.34}$$

and let $T_{\mathcal{B}} : R^k \to R^k$ be the linear transformation defined by $T_{\mathcal{B}}(\vec{c}) = A_{\mathcal{B}}\vec{c}$ for all $\vec{c} \in R^k$. Then for all $\vec{x} \in S$ we have

$$T(\vec{x}) = [A_{\mathcal{B}}[\vec{x}]_{\mathcal{B}}]_{\mathcal{B}}^{-1}. \tag{5.35}$$

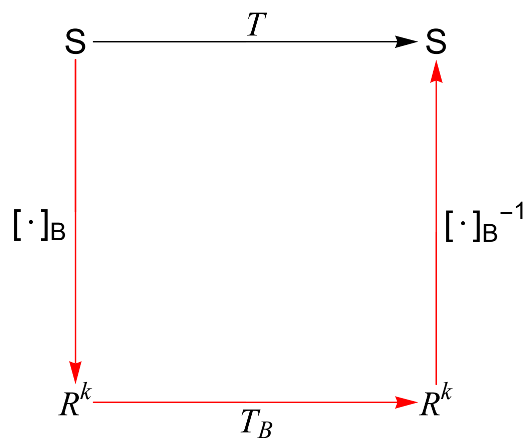


Figure 5.24: $T = [\cdot]_{\mathcal{B}}{}^{-1} \circ T_{\mathcal{B}} \circ [\cdot]_{\mathcal{B}}$

Before giving the formal proof of Theorem 5.6.2, let us try to understand what the theorem is saying in an informal way. Figure 5.24 shows two different routes of getting from S to S. One route (the shorter route indicated in black) is to go directly from S so S using the map T. The other route (longer route indicated in red) is to first go form S to R^k using the coordinate mapping $[\cdot]_{\mathcal{B}}$, then go from R^k to R^k using the map $T_{\mathcal{B}}$, and then go from R^k to S using the inverse coordinate mapping $[\cdot]_{\mathcal{B}}{}^{-1}$.

Figure 5.25 illustrates the effect (using each of the two routes) on a specific vector $\vec{x} \in S$. The direct route takes us directly from $\vec{x}$ to $T(\vec{x})$:

$$\vec{x} \mapsto T(\vec{x}).$$

The indirect route takes us from $\vec{x}$ to $T(\vec{x})$ via the steps

$$\vec{x} \overset{[\cdot]_{\mathcal{B}}}{\mapsto} [\vec{x}]_{\mathcal{B}} \overset{T_{\mathcal{B}}}{\mapsto} A_{\mathcal{B}} [\vec{x}]_{\mathcal{B}} \overset{[\cdot]_{\mathcal{B}}^{-1}}{\mapsto} [A_{\mathcal{B}} [\vec{x}]_{\mathcal{B}}]_{\mathcal{B}}^{-1} = T(\vec{x}).$$

Of course, the reason we need to prove the theorem is to make sure that the equality at the end of the above sequence of mappings is correct. You may have noticed that the notation being used is a bit cumbersome due to the fact that we need to write the subscript "$\mathcal{B}$" so many times. That is true. We really don't need to write it because we are only dealing with one ordered basis $\mathcal{B}$, so there is no room for confusion if we leave the subscript $\mathcal{B}$ out of the notation. Thus let us suppress the writing of $\mathcal{B}$. To economize on notation, we will just write $[\cdot]$ instead of $[\cdot]_B$ with the understanding that

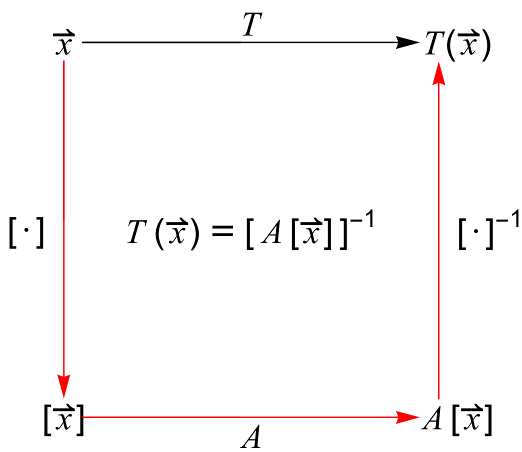


Figure 5.25: $T(\vec{x}) = [A[\vec{x}]]^{-1}$

$[\cdot]$ really means $[\cdot]_{\mathcal{B}}$. We will also suppress the writing of the subscript $\mathcal{B}$ for the matrix $A_{\mathcal{B}}$, thus simply writing A instead of $A_{\mathcal{B}}$. With these notational conveniences, equation (5.35) of Theorem 5.6.2 can be written more simply (with less messy notation) as $T(\vec{x}) = [A[\vec{x}]]^{-1}$. The diagram in Figure 5.24 uses the notation with subscript $\mathcal{B}$ included and the diagram in Figure 5.25 leaves out the subscript $\mathcal{B}$.

And now for the proof of Theorem 5.6.2:

Proof. We want to prove that if $\vec{x}$ is any vector in S, then $T(\vec{x}) = [A[\vec{x}]]^{-1}$. Let $\vec{x} \in S$. Since $\mathcal{B} = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is an ordered basis for S, then there are unique scalars $c_1, c_2, \ldots, c_k$ such that

$$\vec{x} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k.$$

This means that

$$[\vec{x}] = \langle c_1, c_2, \ldots, c_k \rangle .$$

The matrix $A_{\mathcal{B}}$ (which we have decided to just call A) is the matrix whose column vectors are as defined in equation (5.34):

$$\text{Col}_j(A) = [T(\vec{u}_j)] .$$

Recall that $A[\vec{x}]$ is the linear combination of the column vectors of A using the entries of $[\vec{x}]$ as weights, and thus

$$\begin{aligned} A[\vec{x}] &= c_1 \text{Col}_1(A) + c_2 \text{Col}_2(A) + \cdots + c_k \text{Col}_k(A) \\ &= c_1 [T(\vec{u}_1)] + c_2 [T(\vec{u}_2)] + \cdots + c_k [T(\vec{u}_k)] . \end{aligned}$$

Using the coordinate mapping property (5.33), we obtain

$$A[\vec{x}] = [c_1 T(\vec{u}_1)] + [c_2 T(\vec{u}_2)] + \cdots + [c_k T(\vec{u}_k)]$$

and then using property (5.32) we obtain

$$A[\vec{x}] = [c_1 T(\vec{u}_1) + c_2 T(\vec{u}_2) + \cdots + c_k T(\vec{u}_k)].$$

Now we use the fact that T is a linear transformation to obtain

$$\begin{aligned} A[\vec{x}] &= [c_1 T(\vec{u}_1) + c_2 T(\vec{u}_2) + \cdots + c_k T(\vec{u}_k)] \\ &= [T(c_1\vec{u}_1) + T(c_2\vec{u}_2) + \cdots + T(c_k\vec{u}_k)] \\ &= [T(c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k)] \\ &= [T(\vec{x})]. \end{aligned}$$

Since $[T(\vec{x})] = A[\vec{x}]$, then $T(\vec{x}) = [A[\vec{x}]]^{-1}$. □

As is always the case, understanding new ideas requires looking at examples, which we will now do. Before looking at the examples, we provide a theorem that tells us some of the important information we can learn about T by using the matrix A $(= A_{\mathcal{B}})$. We will omit the proof of the theorem.

Theorem 5.6.3. *Suppose that S is a finite dimensional subspace of a vector space V and suppose that $T : S \to S$ is a linear transformation. Suppose also that $\mathcal{B} = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is an ordered basis for S and let A be the matrix of T with respect to the ordered basis $\mathcal{B}$. This is the matrix whose column vectors are given by (5.34). Then*

1. *The set of vectors $\{\vec{y}_1, \vec{y}_2, \ldots, \vec{y}_p\}$ is a basis for* $\text{Range}(T)$ *if and only if the set of vectors $\{[\vec{y}_1], [\vec{y}_2], \ldots, [\vec{y}_p]\}$ is a basis for $\mathcal{CS}(A)$.*

2. *The set of vectors $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_q\}$ is a basis for* $\ker(T)$ *if and only if the set of vectors $\{[\vec{x}_1], [\vec{x}_2], \ldots, [\vec{x}_q]\}$ is a basis for $\mathcal{N}(A)$.*

3. *For any integer $n \geq 1$ and any $\vec{x} \in S$, $T^n(\vec{x}) = [A^n[\vec{x}]]^{-1}$.*

4. *For any vector $\vec{y} \in \text{Range}(T)$, $T(\vec{x}) = \vec{y}$ if and only if $A[\vec{x}] = [\vec{y}]$.*

Part 1 of Theorem 5.6.3 says that to find a basis for $\text{Range}(T)$, we just need to find a basis for $\mathcal{CS}(A)$ and then apply the inverse coordinate mapping to these vectors. Likewise, part 2 of the theorem says that to find a basis for $\ker(T)$, we just need to find a basis for $\mathcal{N}(A)$ and then take the inverse coordinate mapping. Finding bases for $\mathcal{CS}(A)$ and $\mathcal{N}(A)$ is something we know how to do. Part 3 of the theorem is very useful because it allows us to form a composition of T with itself as many times as we like, say n times, by computing A^n. This is something that can be easily done using a calculator (or some other software) even when the matrix A is rather large in size and/or n is large. Part 4 of the theorem tells us that solving an equation of the form $T(\vec{x}) = \vec{y}$ in S is equivalent to solving a matrix–vector equation in R^k.

Example 5.6.9. *Let P_2 be the vector space of polynomial functions of degree 2 or less. We know that $\mathcal{B} = \{1, x, x^2\}$ is an ordered basis for P_2. Let $D : P_2 \to P_2$ be the differentiation transformation defined by $D(p(x)) = p'(x)$.*

Let us answer four questions about D that correspond to the four parts of Theorem 5.6.3. The questions we will answer are

1. *Find* $\text{Range}(D)$.
2. *Find* $\ker(D)$.
3. *Find D^n for all $n \geq 1$.*
4. *Find all functions $p(x) \in P_2$ such that $D(p(x)) = -3 + 2x$.*

We will answer these questions directly (without using either Theorem 5.6.2 or Theorem 5.6.3) and then we will answer them using the theorems. It will be seen that the questions can easily be answered directly, and hence that the theorems are not really needed. Upcoming examples will be ones in which similar questions cannot be as easily answered directly but can be answered using the theorems.

Answering the Questions Directly: *Every element $p(x) \in P_2$ has the form*

$$p(x) = a_0 + a_1 x + a_2 x^2.$$

1) *Using what we learned in calculus about differentiating polynomials, we see that*

$$D(p(x)) = p'(x) = a_1 + 2a_2 x.$$

This tells us that the derivative of a polynomial of degree 2 or less is a polynomial of degree 1 or less and hence that $\text{Range}(D) \subseteq P_1$. *Conversely, if we* *take any polynomial* $q(x) = b_0 + b_1 x$ *in* P_1*, then* $D(p(x)) = q(x)$ *where* $p(x)$ *is the polynomial*

$$p(x) = b_0 x + \frac{1}{2} b_1 x^2.$$

This tells us that $P_1 \subseteq \text{Range}(D)$*. We conclude that* $\text{Range}(D) = P_1$*.*

2) *To find* $\ker(D)$*, we need to find all polynomials* $p(x) = a_0 + a_1 x + a_2 x^2 \in P_2$ *such that* $D(p(x)) = z(x)$ *(where* z *is the zero polynomial). Setting*

$$a_1(1) + 2a_2 x = 0 \text{ for all } x \in R$$

and using the fact that the set of functions $\{1, x\}$ *is linearly independent, we conclude that* $a_1 = 0$ *and* $2a_2 = 0$ *(and hence* $a_2 = 0$*). This tells us that* $\ker(D)$ *contains only constant functions,* $p(x) = a_0$*. Every constant function is a scalar multiple of the constant function* 1*. So we can say that* $\ker(D) = \text{Span}\{1\}$*.*

Notice that we have found that $\text{Range}(D) = P_1 = \text{Span}\{1, x\}$ *and* $\ker(D) = \text{Span}\{1\}$*. We see that* $\text{Range}(D)$ *has dimension* 2 *and* $\ker(D)$ *has dimension* 1 *and that the Fundamental Theorem of Linear Algebra holds true because*

$$\dim(\text{Range}(D)) + \dim(\ker(D)) = 2 + 1 = 3 = \dim(P_2).$$

3) *For a polynomial* $p(x) = a_0 + a_1 x + a_2 x^2 \in P_2$*, we have*

$$\begin{aligned} D(p(x)) &= p'(x) = a_1 + 2a_2 x \\ D^2(p(x)) &= p''(x) = 2a_2 \\ D^3(p(x)) &= p'''(x) = 0 \end{aligned}$$

and $D^n(p(x)) = 0$ *for all* $n \geq 3$*.*

4) *We want to find all solutions of the equation* $D(p(x)) = -3 + 2x$ *that lie in* P_2*. In other words, we want to find all functions* $p(x) \in P_2$ *such that* $p(x)$ *is an* ***antiderivative*** *of* $-3 + 2x$*. In calculus, we learned that*

$$\int (-3 + 2x)\, dx = C - 3x + x^2$$

where C *can be any constant. Hence the solutions of* $D(p(x)) = -3 + 2x$ *that lie in* P_2 *are all functions of the form* $p(x) = C - 3x + x^2$*. There are infinitely many solutions because* C *can be any constant.*

Answering the Questions using the Theorems: *First we will use Theorem 5.6.2 to construct the matrix of D with respect to the ordered basis $\mathcal{B} = \{1, x, x^2\}$. Since*

$$\begin{aligned} D(1) &= 0 \\ D(x) &= 1 \\ D\left(x^2\right) &= 2x, \end{aligned}$$

then

$$\begin{aligned} [D(1)] &= \langle 0,0,0 \rangle \\ [D(x)] &= \langle 1,0,0 \rangle \\ \left[D\left(x^2\right)\right] &= \langle 0,2,0 \rangle . \end{aligned}$$

Thus we have by Theorem 5.6.2 that

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

1) *Since*

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \operatorname{rref}(A), \tag{5.36}$$

we see that the set of vectors $\{\langle 1,0,0 \rangle, \langle 0,2,0 \rangle\}$ is a basis for $\mathcal{CS}(A)$ (because these are the pivot columns of A). Theorem 5.6.3 then tells us that the set of vectors

$$\left\{[\langle 1,0,0 \rangle]^{-1}, [\langle 0,2,0 \rangle]^{-1}\right\} = \{1, 2x\}$$

is a basis for $\operatorname{Range}(D)$. *Since* $\operatorname{Span}\{1, 2x\}$ *is the same thing as* $\operatorname{Span}\{1, x\}$, *we can say that*

$$\operatorname{Range}(D) = \operatorname{Span}\{1, x\} = P_1.$$

2) *The row reduction done in (5.36) also shows us that a basis for $\mathcal{N}(A)$ is $\{\langle 1,0,0 \rangle\}$ and Theorem 5.6.3 then tells us that a basis for* $\ker(D)$ *is*

$$\left\{[\langle 1,0,0 \rangle]^{-1}\right\} = \{1\}.$$

Hence $\ker(D) = \operatorname{Span}\{1\}$, *which is the set of all constant functions.*

3) Note that

$$A^2 = \left(\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}\right)^2 = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$A^3 = \left(\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}\right)^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = O_{3\times 3}$$

and hence $A^n = O_{3\times 3}$ *for all* $n \geq 3$.

Now observe that for any polynomial $p(x) = a_0 + a_1 x + a_2 x^2$ *in* P_2 *we have*

$$A[p(x)] = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \langle a_0, a_1, a_2 \rangle = \langle a_1, 2a_2, 0 \rangle$$

and thus part of 3 of Theorem 5.6.3 tells us that

$$D(p(x)) = [A[p(x)]]^{-1} = a_1 + 2a_2 x.$$

Likewise

$$A^2[p(x)] = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \langle a_0, a_1, a_2 \rangle = \langle 2a_2, 0, 0 \rangle$$

and hence

$$D^2(p(x)) = 2a_3.$$

For any $n \geq 3$ *we have*

$$A^n[p(x)] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \langle a_0, a_1, a_2 \rangle = \langle 0, 0, 0 \rangle$$

and hence

$$D^n(p(x)) = 0 \text{ for all } n \geq 3.$$

4) To solve $D(p(x)) = -3 + 2x$*, we can solve the matrix–vector equation* $A[p(x)] = [-3 + 2x]$ *and then apply the inverse coordinate map. Since* $[-3 + 2x] = \langle -3, 2, 0 \rangle$*, then to solve* $A[p(x)] = [-3 + 2x]$*, we need to perform row reduction on the augmented matrix*

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

The row reduction can be achieved in just one step (a scaling operation on the second row). We obtain

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

If we call the unknowns of our equation $\langle a_0, a_1, a_2 \rangle$ *then we have*

$$\begin{aligned} a_0 &= C \ \textit{(a free variable)} \\ a_1 &= -3 \\ a_2 &= 1 \end{aligned}$$

which we can write in the vector form $\langle a_0, a_1, a_2 \rangle = \langle C, -3, 1 \rangle$. *Part 4 of Theorem 5.6.3 then tells us that the solution set of* $D(p(x)) = -3 + 2x$ *consists of all functions of the form* $p(x) = C - 3x + x^2$ *where* C *can be any constant.*

Example 5.6.10. *Let*

$$S = \text{Span}\left\{\sin(x), \cos(x), x\sin(x), x\cos(x)\right\}.$$

We leave it as an exercise to show that the set of vectors

$$\mathcal{B} = \left\{\sin(x), \cos(x), x\sin(x), x\cos(x)\right\}$$

is linearly independent and is thus an ordered basis for S. *We once again consider the differentiation transformation* $D : S \to S$. *To find the matrix of* D *with respect to the ordered basis* $\mathcal{B}$, *we compute*

$$\begin{aligned} D(\sin(x)) &= \cos(x) \\ D(\cos(x)) &= -\sin(x) \\ D(x\sin(x)) &= \sin(x) + x\cos(x) \\ D(x\cos(x)) &= -x\sin(x) + \cos(x). \end{aligned} \tag{5.37}$$

and then observe that

$$\begin{aligned} [D(\sin(x))] &= \langle 0, 1, 0, 0 \rangle \\ [D(\cos(x))] &= \langle -1, 0, 0, 0 \rangle \\ [D(x\sin(x))] &= \langle 1, 0, 0, 1 \rangle \\ [D(x\cos(x))] &= \langle 0, 1, -1, 0 \rangle. \end{aligned}$$

From this we obtain the matrix of D with respect to the basis $\mathcal{B}$, which is

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

Since

$$\text{rref}\,(A) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

then

$$\begin{aligned} \mathcal{CS}\,(A) &= R^4 \\ \mathcal{N}\,(A) &= \left\{\vec{0}_4\right\} \end{aligned}$$

and Theorem 5.6.3 then tells us that

$$\begin{aligned} \text{Range}\,(D) &= S \\ \ker\,(D) &= \{z\}\,. \end{aligned}$$

(Recall that we are using the notation z to denote the zero function.)

To illustrate part 3 of Theorem 5.6.3, suppose that we wish to compute the fifth derivative of the function $f(x) = 3x\sin(x)$. We know how to do this using the Product Rule of differentiation that we learned in Calculus I, but it is tedious. We would start by computing the first derivative

$$D\,(3x\sin(x)) = 3x\cos(x) + 3\sin(x)\,.$$

Then we would compute the second derivative

$$\begin{aligned} D^2\,(3x\sin(x)) &= D\,(3x\cos(x) + 3\sin(x)) \\ &= 3x\,(-\sin(x)) + 3\cos(x) + 3\cos(x) \\ &= -3x\sin(x) + 6\cos(x)\,, \end{aligned}$$

and then we would need to repeat this process three more times to obtain the fifth derivative. However, computing the fifth derivative can easily be achieved *using the matrix*

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

(and a calculator) to obtain

$$A^5 = \left(\begin{bmatrix} 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \right)^5 = \begin{bmatrix} 0 & -1 & 5 & 0 \\ 1 & 0 & 0 & 5 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

The matrix A^5 *is the matrix of* D^5 *with respect to the ordered basis* $\mathcal{B}$*. Since* D^5 *is the fifth derivative transformation, then to find* $D^5(3x\sin(x))$*, we can instead find* $A^5\langle 0,0,3,0\rangle$ *because* $\langle 0,0,3,0\rangle$ *is the coordinate vector of* $3x\sin(x)$ *with respect to the ordered basis* $\mathcal{B}$*. After doing that, we convert the answer back to the "S world" by applying the inverse coordinate mapping.*

Since

$$A^5\langle 0,0,3,0\rangle = \begin{bmatrix} 0 & -1 & 5 & 0 \\ 1 & 0 & 0 & 5 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \langle 0,0,3,0\rangle = \langle 15,0,0,3\rangle,$$

then the inverse coordinate mapping gives

$$D^5(3x\sin(x)) = 15\sin(x) + 3x\cos(x).$$

We can be completely general. Suppose we start with any function, f*, in* S *and suppose we want to compute the* n*th derivative of this function. That is, we want to compute* $D^n(f)$*. Since* f *has the form*

$$f(x) = c_1\sin(x) + c_2\cos(x) + c_3x\sin(x) + c_4x\cos(x),$$

then

$$[f(x)] = \langle c_1, c_2, c_2, c_4\rangle.$$

To find $D^n(f(x))$*, we just compute* $A^n[f(x)]$ *and then apply the inverse coordinate map to obtain* $D^n(f(x))$*.*

As an example, suppose we want to find the tenth derivative of the function

$$f(x) = \cos(x) + x\sin(x) + 2x\cos(x).$$

This means that we want to find $D^{10}(f(x))$*. Since the matrix of* D^{10} *with respect to the ordered basis* $\mathcal{B}$ *is*

$$A^{10} = \left(\begin{bmatrix} 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}\right)^{10} = \begin{bmatrix} -1 & 0 & 0 & -10 \\ 0 & -1 & 10 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix},$$

and since

$$[f(x)] = \langle 0, 1, 1, 2 \rangle,$$

and since

$$A^{10}[f(x)] = \langle -20, 9, -1, -2 \rangle,$$

then

$$D^{10}(\cos(x) + x\sin(x) + 2x\cos(x)) = -20\sin(x) + 9\cos(x) - x\sin(x) - 2x\cos(x).$$

*(**The continuation of this example may be most appreciated by students who have studied integration techniques in Calculus II. However, the continuation of the example really only requires an understanding of what an antiderivative is.**)*

As an illustration of part 4 of Theorem 5.6.3, we can also use the matrix A *to compute integrals of the form*

$$\int f(x)\, dx$$

where f *is a function in the vector space* S*. For example, suppose we wish to compute the integral*

$$\int x\sin(x)\, dx.$$

If you have studied integration techniques in Calculus II, then you probably remember how to do this problem: use integration by parts. However, since this is an antidifferentiation problem, we realize that what we are doing is looking for the solutions of the equation $D(F(x)) = x\sin(x)$*. Realizing that the coordinate vector of* $x\sin(x)$ *with respect to the ordered basis*

$\mathcal{B}$ *is* $[x\sin(x)] = \langle 0,0,1,0\rangle$, *we see that the equation that we need to solve is* $A[F(x)] = \langle 0,0,1,0\rangle$. *Forming the appropriate augmented matrix and performing row reduction we obtain*

$$\left[\begin{array}{cccc|c} 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right]$$

and we see that the unique solution of $A[F(x)] = \langle 0,0,1,0\rangle$ *is* $[F(x)] = \langle 1,0,0,-1\rangle$. *When we translate this fact back into the world of* S, *it means that the unique solution,* $F(x) \in S$, *of* $D(F(x)) = x\sin(x)$ *is* $F(x) = \sin(x) - x\cos(x)$. *Thus*

$$\int x\sin(x)\ dx = \sin(x) - x\cos(x).$$

You might be thinking that the above answer can't be right. In fact, it is not quite right. The equation $D(F) = x\sin(x)$ *actually has infinitely many solutions. The correct answer is*

$$\int x\sin(x)\ dx = \sin(x) - x\cos(x) + C$$

where C *can be any constant. Why did our method not yield the "+C" that should be there? The reason is that we started by considering the vector space*

$$S = \text{Span}\{\sin(x), \cos(x), x\sin(x), x\cos(x)\}$$

and this space does not include any constant functions other than the zero function $z(x) = 0$. *If we had started, instead, with the vector space*

$$S^* = \text{Span}\{1, \sin(x), \cos(x), x\sin(x), x\cos(x)\},$$

which does include all of the constant functions, then our matrix A *would be a* 4×5 *matrix. One free variable would be present and that free variable gives us the "+C". The problem that we solved was not to find all functions* $F \in S^*$ *such that* $D(F) = x\sin(x)$. *The problem that we solved was to find all functions* $F \in S$ *such that* $D(F) = x\sin(x)$.

Exercise 5.6.15. *For the vector space* $P_3 = \text{Span}\{1, x, x^2, x^3\}$ *with ordered basis* $\mathcal{B} = \{1, x, x^2, x^3\}$, *let* $D : P_3 \to P_3$ *be the differentiation transformation* $D(p) = p'$.

1. *Find the matrix, A, of D with respect to the basis $\mathcal{B}$. .*

2. *Use A to determine* $\operatorname{Range}(D)$ *and* $\ker(D)$.

3. *Compute A^2, A^3, and A^4. You should find that $A^4 = O_{4\times 4}$.*

Exercise 5.6.16. *Let $\mathcal{B} = \{1, e^x, e^{2x}\}$. This set of functions is linearly independent and is hence an ordered basis for $S = \operatorname{Span}\{1, e^x, e^{2x}\}$. Let $D : S \to S$ be the differentiation transformation $D(f) = f'$.*

1. *Using calculus, we obtain*

$$\begin{aligned} D(1) &= \_\_\_\_\_\_\_\_\_ \\ D(e^x) &= \_\_\_\_\_\_\_\_\_ \\ D\left(e^{2x}\right) &= \_\_\_\_\_\_\_\_\_ \end{aligned}$$

2. *Find the matrix, A, of D with respect to the basis $\mathcal{B}$.*

3. *It is not too tedious to compute the third derivative of the function $f(x) = 4 - 6e^x + 3e^{2x}$ using calculus. Please do compute $f'''(x) = D^3(f(x))$ using calculus.*

4. *Use part 2 of Theorem 5.6.3 to compute $f'''(x)$ by using the matrix that you found in question 2.*

Exercise 5.6.17. *Let $\mathcal{B} = \{1, e^x \sin(x), e^x \cos(x)\}$. This set of functions if linearly independent and is hence a basis for $S = \operatorname{Span}\{1, e^x \sin(x), e^x \cos(x)\}$.*

Use an appropriate matrix to find the fifth derivative of the function

$$f(x) = 5 - e^x \sin(x).$$

(It is a bit tedious to do this by just using calculus, but you may want to try it just to see that your answer matches with what you got using a matrix.)

Exercise 5.6.18. *Let $\mathcal{B} = \{1, x, e^x, xe^x\}$. This set of functions is linearly independent and is hence an ordered basis for $S = \operatorname{Span}\{1, x, e^x, xe^x\}$. Let $D : S \to S$ be the differentiation transformation $D(f) = f'$.*

1. *Using calculus, we obtain*

$$\begin{aligned} D(1) &= \_\_\_\_\_\_ \\ D(x) &= \_\_\_\_\_\_ \\ D(e^x) &= \_\_\_\_\_\_ \\ D(xe^x) &= \_\_\_\_\_\_. \end{aligned}$$

2. *Find the matrix, A, of D with respect to the basis $\mathcal{B}$.*

3. *Compute the* 6*th derivative of the function*

$$f(x) = 5 + 4x + 3e^x - 2xe^x$$

using the matrix that you found in question 2.

Exercise 5.6.19. *Use the matrix you found in Exercise 5.6.18 to evaluate the indefinite integral*

$$\int (2 - 2x + 3e^x + 2xe^x)\, dx.$$

Exercise 5.6.20. *(**perhaps best appreciated by students who have had Calculus II, but only requires an understanding of the concept of antiderivative**) You may remember from Calculus II that integrals of the form*

$$\int (ae^x \sin(x) + be^x \cos(x))\, dx$$

can be computed using integration by parts twice, which can be tedious.

Use the matrix you found in Exercise 5.6.17 to evaluate

$$\int e^x \sin(x)\, dx.$$

5.7 Isomorphism of Vector Spaces

In Section 5.6.3 we made heavy use of the coordinate mapping from a finite dimensional vector space, S, of dimension k to the vector space R^k. The coordinate mapping allowed us to work with vectors in R^k in order to draw conclusions about several important properties of the vector space S. It also allowed us to solve equations involving vectors in S by solving the corresponding equations in R^k. A rather informal description of what we learned in looking at many examples is that no matter what the nature of the vectors in S, the fact that S is k–dimensional means that S is "essentially the same" as R^k as far as algebra is concerned. This formal term that corresponds to the informal idea of "essentially the same" is *isomorphic*. This comes from the Ancient Greek *isos* meaning "same" and *morphe* meaning "form" or "shape".

Definition 5.7.1. *A vector space V is said to be* ***isomorphic*** *to a vector space W if there exists an invertible linear transformation $T : V \to W$. Any invertible linear transformation $T : V \to W$ is said to be an* ***isomorphism*** *from V onto W.*

Remark 5.7.1. *Some basic observations concerning Definition 5.7.1 are*

1. *Any vector space, V, is isomorphic to itself because the identity transformation $E : V \to V$ is an isomorphism from V onto V.*
2. *If V is isomorphic to W, then W is isomorphic to V. This is because if $T : V \to W$ is an isomorphism that $T^{-1} : W \to V$ is also an isomorphism.*
3. *If V is isomorphic to W and W is isomorphic to X, then V is isomorphic to X.*

Exercise 5.7.1. *Prove part 3 of Remark 5.7.1.*

Remark 5.7.2. *By part 2 of Remark 5.7.1, a vector space V is isomorphic to a vector space W if and only if W is isomorphic to V. Thus it makes sense, when we know that V is isomorphic to W, to say that V and W are* ***isomorphic to each other****.*

As mentioned above, the type of isomorphism that we have studied thus far is a coordinate mapping. If S is a finite dimensional vector space (which might be a subspace of some larger finite or infinite dimensional vector space, V) and $\mathcal{B} = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is an ordered basis for S consisting of exactly $k \geq 1$ vectors, then $\dim(S) = k$ and the coordinate mapping

$$\vec{x} \overset{[\cdot]_{\mathcal{B}}}{\mapsto} [\vec{x}]_{\mathcal{B}}$$

is an isomorphism from S onto R^k. This is because $[\cdot]_{\mathcal{B}}$ is invertible. Of course it is also true that the inverse coordinate mapping

$$[\vec{x}]_{\mathcal{B}} \overset{[\cdot]_{\mathcal{B}}^{-1}}{\mapsto} \vec{x}$$

is an isomorphism from R^k onto S. Hence S and R^k are isomorphic to each other.

The following theorem tells us that two finite dimensional vector spaces are isomorphic to each other if and only if these vector spaces have the same dimension.

Theorem 5.7.1. *Suppose that V and W are finite–dimensional vector spaces. Then V and W are isomorphic to each other if and only if* $\dim(V) = \dim(W)$*. Specifically, if V and W both have dimension k (where $1 \leq k < \infty$), then V and W are both isomorphic to R^k.*

Proof. We already know that if V is a finite dimensional vector space with $\dim(V) = k$ (where $1 \leq k < \infty$), then V is isomorphic to R^k. The reason is that we can choose any ordered basis $\mathcal{B} = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ for V and we know that the coordinate mapping $[\cdot]_{\mathcal{B}} : V \to R^k$ is an isomorphism from V onto R^k.

Now suppose that V and W both have dimension k. Then V is isomorphic to R^k and R^k is isomorphic to W. By part 3 of Remark 5.7.1, V is isomorphic to W.

Conversely, suppose that V and W are both finite dimensional and that V is isomorphic to W. This means that there exists an isomorphism $T : V \to W$. By the Fundamental Theorem of Linear Algebra,

$$\dim(\text{Range}(T)) + \dim(\ker(T)) = \dim(V).$$

Since T maps V onto W then $\text{Range}(T) = W$ and hence $\dim(\text{Range}(T)) = \dim(W)$.

Since T is one–to–one, then $\ker(T) = \left\{\vec{0}_V\right\}$ and hence $\dim(\ker(T)) = 0$. We now see that

$$\dim(W) + 0 = \dim(V).$$

This completes the proof. □

Example 5.7.1. *The vector space, $M_{2\times2}$, consisting of all 2×2, matrices has dimension 4.*

The vector space, $M_{1\times 4}$, consisting of all 1×4, matrices has dimension 4.

The vector space $P_3 = \text{Span}\{1, x, x^2, x^3\}$, consisting of all polynomial functions that have degree 3 or less, has dimension 4.

The vector space $S = \text{Span}\{1, x, x\sin(x), x\cos(x)\}$ has dimension 4.

All four of the above–mentioned vector spaces have the same dimension (4) and hence all of these vector spaces are isomorphic to each other. Each of them is isomorphic to R^4.

Exercise 5.7.2. *Which of the following pairs of vector spaces are isomorphic to each other? (This is equivalent to asking whether or not the given pair of vector spaces have the same dimension.)*

1. *R^5 and R^6*
2. *R^5 and P_4*
3. *$\text{Span}\{1, x, x^2\}$ and $\text{Span}\{1, e^x, e^{2x}\}$*
4. *$\left\{\vec{0}_2\right\}$ and $\left\{\vec{0}_5\right\}$*
5. *A line through the origin in R^2 and a line through the origin in R^3*

5.8 Additional Exercises

(Jump to Solutions)

1. The **identity transformation** $E : R^2 \to R^2$ is defined by

$$E(\vec{x}) = \vec{x}.$$

   (a) Show that E satisfies both of the requirements of Definition 5.2.1 and is thus a linear transformation.

   (b) Suppose that L is any line in R^2. To what line does the identity transformation map L? In other words, what is $E(L)$?

2. The **zero transformation** $Z : R^2 \to R^2$ is defined by

$$Z(\vec{x}) = \vec{0}_2.$$

(a) Show that Z satisfies both of the requirements of Definition 5.2.1 and is thus a linear transformation.

(b) Suppose that L is any line in R^2. What is $Z(L)$?

3. We have stated that a function that satisfies the linearity properties of Definition 5.2.1 is called a linear function and that a function that does not satisfy those properties is called a nonlinear function. According to this, a linear function $f : R \to R$ is a function that satisfies the properties

   1. If x and y are any two real numbers, then $f(x+y) = f(x)+f(y)$.
   2. If x is any real number and c is a scalar (a real number), then $f(cx) = cf(x)$.

   This exercise points out a nuance in this terminology.

   a) Show that the function $f : R \to R$ defined by $f(x) = 5x$ is a linear function.

   b) Show that if a is any constant (real number), then the function $f : R \to R$ defined by $f(x) = ax$ is a linear function.

   c) Explain why the function $f : R \to R$ defined by $f(x) = 5x + 3$ is not a linear function.

   You may be surprised and disappointed to learn that $f(x) = 5x + 3$ is not a linear function, because its graph is a straight line and you were probably taught throughout all of the math courses you have taken (including calculus) that this kind of function is called a linear function. It is actually called an **affine function**. In linear algebra, a function $f : R \to R$ that has the form $f(x) = ax + b$, where a and b are constants, is called a linear function only if $b = 0$. If $b \neq 0$, then f is called an affine function. When $b \neq 0$, we don't want to call the function linear because it does not satisfy the linearity requirements of Definition 5.2.1. In other courses, such as calculus, there is a different interpretation of the word "linear". It just means a function whose graph is a straight line. It would be distracting in studying calculus if we were to make a distinction between linear and affine functions. so we don't do it in calculus. However, it is essential to make this distinction in linear algebra.

4. For the following linear transformations $T : R^n \to R^m$, determine the range of T and the kernel of T. Also determine whether or not T is invertible. If T is invertible, then find the formula for T^{-1}.

   (a) $T : R^2 \to R^2$ defined by $T(\langle x_1, x_2\rangle) = \langle -4x_1 + 2x_2, -4x_1 - 6x_2\rangle$
   (b) $T : R^2 \to R^2$ defined by $T(\langle x_1, x_2\rangle) = \langle -4x_1 + 2x_2, -4x_1 + 2x_2\rangle$
   (c) $T : R^2 \to R^4$ defined by $T(\langle x_1, x_2\rangle) = \langle x_1, x_1, x_2, x_1 + x_2\rangle$
   (d) $T : R^5 \to R^2$ defined by $T(\langle x_1, x_2, x_3, x_4, x_5\rangle) = \langle 5x_1 - 3x_2 - 3x_3 - 5x_4 - 2x_5, x_3\rangle$
   (e) $T : R^3 \to R^3$ defined by $T(\langle x_1, x_2, x_3\rangle) = \langle x_1, 0, 0\rangle$

5. Find the linear transformation that reflects vectors in R^2 through the x_1 axis. To do this

   (a) Determine $T(\langle 1, 0\rangle)$ and $T(\langle 0, 1\rangle)$.
   (b) Use what you found in part a to write down the matrix A such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in R^2$.
   (c) Write the formula for T in the form $T(\langle x_1, x_2\rangle) = \langle \_\_\_\_, \_\_\_\_\rangle$.

6. Find the linear transformation that reflects vectors in R^2 through the x_2 axis. You can do this by following the same procedure as in Exercise 5.

7. Let L be the line, pictured in Figure 5.26, that makes an angle of θ with the positive x_1 axis.

   The purpose of this exercise is to find the linear transformation that reflects vectors through L. Call this linear transformation T. Thus

   $$T(\vec{x}) = \text{ reflection of } \vec{x} \text{ through the line } L.$$

   Find the standard matrix, A, of T and also write T in the form

   $$T(\langle x_1, x_2\rangle) = \langle \_\_\_\_\_, \_\_\_\_\_\rangle.$$

   Here is the suggested strategy:

   In Section 5.3.6, we determined the linear transformation, R_θ, that rotates vectors by some angle θ.

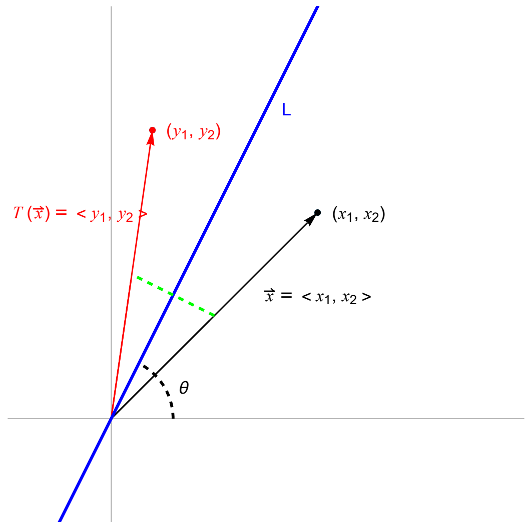


Figure 5.26: $T(\vec{x})$ = reflection of $\vec{x}$ through L.

In Exercise 5 you found the linear transformation that reflects vectors through the x_1 axis. Give this linear transformation the name S.

Observe that reflecting a vector $\vec{x}$ through the line L can be accomplished in three steps:

(a) First apply $R_{-\theta}$, which is the same thing as R_θ^{-1}, to $\vec{x}$. That will rotate $\vec{x}$ through the angle $-\theta$.

(b) Then reflect $R_\theta^{-1}(\vec{x})$ through the x_1 axis by applying the linear transformation S to it.

(c) Then rotate $\left(S \circ R_\theta^{-1}\right)(\vec{x})$ back through the angle θ by applying R_θ to it.

At some point in doing this exercise you may find it helpful to remember the trigonometric identities

$$\cos^2(\theta) - \sin^2(\theta) = \cos(2\theta)$$
$$2\sin(\theta)\cos(\theta) = \sin(2\theta).$$

8. Use the general result that you found in Exercise 7 to find the linear transformation $T : R^2 \to R^2$ that reflects vectors through the following lines L:

(a) L is the line that makes an angle of 60° with the positive x_1 axis.

(b) L is the line $x_2 = 2x_1$.

9. Show that the reflection transformation that you found in Exercise 7 is its own inverse. Does this make sense?

10. Two $n \times n$ matrices, A and B, are similar to each other if there exists an invertible $n \times n$ matrix C such that $A = C^{-1}BC$. If we multiply both sides of this equation on the left by C, we obtain

$$CA = C\left(C^{-1}BC\right)$$

which can write as

$$CA = \left(CC^{-1}\right)(BC)$$

or as

$$CA = I_n BC$$

or as

$$CA = BC.$$

Thus A and B are similar to each other if there exists an invertible $n \times n$ matrix C such that $CA = BC$.

Show that the two matrices

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

are similar to each other by finding an invertible 2×2 matrix C such that $CA = BC$.

Hint to set this problem up: Let $C = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$, then plug this into $CA = BC$ and solve for C.

11. (**for students who have studied infinite sequences in Calculus II**) In Section 4.7.3, it was pointed out that the set of all convergent sequences of real numbers,

$$C = \{\vec{a} = \langle a_1, a_2, a_3, \ldots \rangle \in R^\infty \mid \vec{a} \text{ converges}\}$$

is a subspace of R^∞. Recall that "$\vec{a}$ converges" means that there exists a real number $L_{\vec{a}}$ such that

$$\lim_{n\to\infty} a_n = L_{\vec{a}}.$$

The number $L_{\vec{a}}$ is called the **limit** of the sequence $\vec{a} = \langle a_1, a_2, a_3, \ldots\rangle$. As a specific example, suppose that $\vec{a}$ is the sequence $\vec{a} = \left\langle 1, \frac{1}{2}, \frac{1}{3}, \ldots \right\rangle$, meaning that $a_n = \frac{1}{n}$ for each $n = 1, 2, 3, \ldots$. Then $\vec{a}$ is in the subspace C because $\vec{a}$ converges. The limit of $\vec{a}$ is

$$\lim_{n\to\infty} \frac{1}{n} = 0 = L_{\vec{a}}.$$

Consider the function $T : C \to R$ defined by $T(\vec{a}) = L_{\vec{a}}$ where $L_{\vec{a}}$ is the limit of the sequence $\vec{a}$.

(a) Verify that T satisfies both requirement of Definition 5.6.1 and is thus a linear transformation.

(b) Describe $\text{Range}(T)$ and $\ker(T)$.

(c) Does T map C onto R? Explain.

(d) Is T one–to–one? Explain.

(e) Is T invertible? Explain.

12. The linear transformation $S : R^2 \to R^2$ defined by

$$S(\langle x_1, x_2\rangle) = \langle 3x_1, x_2\rangle$$

multiplies that x_1 component of each vector by 3 and leaves the x_2 component unchanged.

The linear transformation $T : R^2 \to R^2$ defined by

$$T(\langle x_1, x_2\rangle) = \left\langle x_1, \frac{1}{2}x_2 \right\rangle$$

multiplies that x_2 component of each vector by $\frac{1}{2}$ and leaves the x_1 component unchanged.

(a) Explain in words what $T \circ S$ does to vectors in R^2.

(b) Write down the matrices A_S and A_T and $A_{T\circ S}$.

(c) Write a formula for $T \circ S$ in the form $(T \circ S)(\langle x_1, x_2\rangle) = $ ______.

(d) Show that $S \circ T = T \circ S$.

(e) For the rectangle pictured in Figure 10 , draw the image of the rectangle under $T \circ S$.

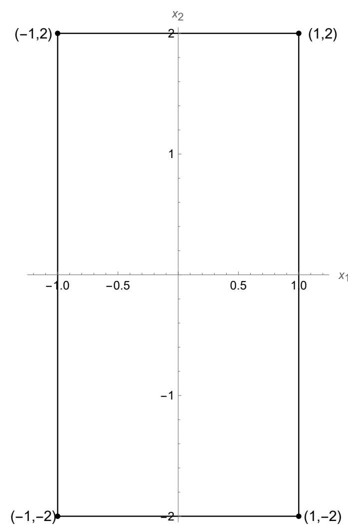


Figure 5.27: Rectangle for Question 12

13. Let $\mathcal{B} = \{\sin(x), \cos(x)\}$. Since $\mathcal{B}$ is linearly independent, it is a basis for $S = \text{Span}\{\sin(x), \cos(x)\}$. Let $D : S \to S$ be the differentiation transformation $D(f) = f'$.

    (a) Find the matrix, A, of D with respect to the ordered basis $\mathcal{B}$.

    (b) Compute A^2, A^3, and A^4.

    (c) Use the matrices A, A^2, A^3, and A^4 to compute the first four derivatives of $f(x) = \sin(x)$.

14. Let $\mathcal{B} = \{e^x, e^{-x}\}$. Since $\mathcal{B}$ is linearly independent, it is a basis for $S = \text{Span}\{e^x, e^{-x}\}$. Let $D : S \to S$ be the differentiation transformation $D(f) = f'$.

    (a) Find the matrix, A, of D with respect to the ordered basis $\mathcal{B}$.

(b) Use the matrix A to compute the 23rd derivative of the function $f(x) = -e^x + 2e^{-x}$.

Chapter 6

Eigenstuff

In Chapter 5, we saw that a linear transformation that maps a vector in R^n to a vector in R^m can always be characterized by a matrix-vector product, $A\vec{x}$. Specifically, a linear transformation $T : R^n \to R^m$ will satisfy $T(\vec{x}) = A\vec{x}$, where A is an $m \times n$ matrix. In this chapter, we will continue to consider linear transformations with our focus restricted to the case in which the vector $\vec{x}$ and $T(\vec{x})$ are in the same vector space R^n. Of course, this means that such a transformation, T, maps R^n into R^n, and the corresponding matrix will be a square matrix.

Consider the linear transformation $T : R^2 \to R^2$ defined by

$$T(\langle x_1, x_2\rangle) = \langle 5x_1 - x_2, 3x_1 + x_2\rangle.$$

Let $\vec{v} = \langle 1, 3\rangle$ and $\vec{u} = \langle -1, 1\rangle$. In Figure 6.1, we see the standard representations of the vectors $\vec{v}$ and $\vec{u}$ together with their images

$$T(\vec{v}) = \langle 2, 6\rangle, \quad \text{and} \quad T(\vec{u}) = \langle -6, -2\rangle.$$

We see that the effect of the transformation T on the vector $\vec{u}$ has two properties. It appears to rotate the vector $\vec{u}$ through a counterclockwise angle (i.e., change its direction) and to increase the magnitude. When comparing $\vec{v}$ and its image $T(\vec{v})$, we see a change in magnitude, but the direction is not changed. In fact, we can say that $T(\vec{v})$ is an element of Span$\{\vec{v}\}$, making $T(\vec{v}) = \lambda\vec{v}$ for some real number λ. (It's easy to see that $\lambda = 2$ in this case, since $\langle 2, 6\rangle = 2\langle 1, 3\rangle$.) The vector $\vec{v}$ seems to have a special relationship with this linear transformation T and its standard matrix, $A = \begin{bmatrix} 5 & -1 \\ 3 & 1 \end{bmatrix}$, since

$T(\vec{v}) = A\vec{v}$. If we consider any vector $\vec{x}$ in $\text{Span}\{\vec{v}\}$, then $\vec{x} = c\vec{v}$ for some scalar c. By the algebraic properties of the matrix-vector product, we see that

$$A\vec{x} = A(c\vec{v}) = cA\vec{v} = c(2\vec{v}) = 2(c\vec{v}) = 2\vec{x}.$$

So the matrix A scales every vector in $\text{Span}\{\vec{v}\}$ by this same factor of 2. Given the complex nature of the matrix-vector product, this is an interesting observation. A relationship such as $A\vec{x} = 2\vec{x}$ is different from equations we've encountered in previous chapters. Most notably, the product on the left, $A\vec{x}$, is a matrix-vector product while the product on the right, $2\vec{x}$, is scalar multiplication of a vector. In general, these are not comparable products. We can rephrase the scalar multiplication $2\vec{x}$ in terms of a matrix-vector product if we make use of the identity matrix I_2. Given that $I_2\vec{x} = \vec{x}$, we can write $2\vec{x} = 2I_2\vec{x}$. Then for any vector $\vec{x}$ in $\text{Span}\{\langle 1, 3\rangle\}$, we have

$$A\vec{x} = 2I_2\vec{x},$$

which can be rearranged into a homogeneous matrix-vector equation

$$\big(A - 2I_2\big)\vec{x} = \vec{0}_2. \tag{6.1}$$

Note that

$$A - 2I_2 = \begin{bmatrix} 5 & -1 \\ 3 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 3 & -1 \\ 3 & -1 \end{bmatrix}.$$

With equivalent rows (making them linearly dependent), it's immediately apparent that this matrix is not invertible. In fact,

$$\text{rref}(A - 2I_2) = \begin{bmatrix} 1 & -1/3 \\ 0 & 0 \end{bmatrix} \neq I_2.$$

This is consistent with the fact that we already know that there are nontrivial solutions to the homogeneous Equation (6.1). In fact, from this rref, we see that solutions to Equation (6.1) will be of the form $\vec{x} = t\,\langle 1/3, 1\rangle$ for $t \in R$. The vector $\vec{v}$ that we started with is of this form (with the choice $t = 3$).

You may be thinking that this example is contrived, and that's a fair observation. Once we know that there is a nonzero vector $\vec{v}$ for which $A\vec{v} = 2\vec{v}$, we can use the tools we have developed to identify these vectors. But it's not clear where that equation came from (or why the scalar 2 features in it). The example raises a number of questions. For example, are there other vectors, not in $\text{Span}\{\vec{v}\}$, such that the product $A\vec{x}$ reduces to scalar multiplication?

Are there other scalars, besides 2, that hold a similar relationship to the matrix A? We will answer questions of this sort in this chapter as well as explore some of the advantages of such special linear transformation-matrix-vector relationships.

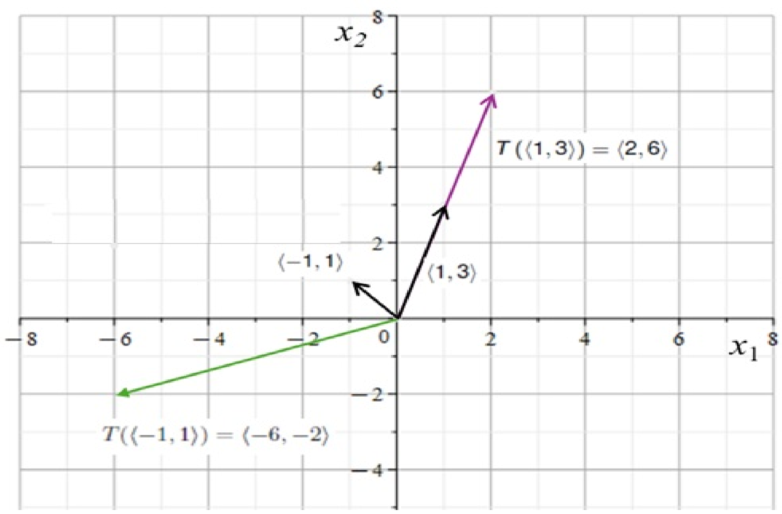


Figure 6.1: Standard representations of vectors $\vec{u} = \langle -1, 1\rangle$ and $\vec{v} = \langle 1, 3\rangle$ along with their images under $T(\langle x_1, x_2\rangle) = \langle 5x_1 - x_2, 3x_1 + x_2\rangle$.

Exercise 6.0.1. *For the matrix* $A = \begin{bmatrix} 5 & -1 \\ 3 & 1 \end{bmatrix}$,

1. *Evaluate* $A\vec{x}$ *where* $\vec{x} = \langle 1, 1\rangle$.

2. *Show that if* $\vec{x}$ *is any vector in* $\mathrm{Span}\{\langle 1, 1\rangle\}$, *then* $A\vec{x} = 4\vec{x}$.

3. *Identify the matrix* $A - 4I_2$, *and show that this matrix is not invertible.*

Exercise 6.0.2. *Consider the matrix* $A = \begin{bmatrix} 4 & 7 \\ 2 & -1 \end{bmatrix}$.

1. *Find a nonzero vector* $\vec{v} = \langle v_1, v_2\rangle$ *such that* $A\vec{v} = 6\vec{v}$.

2. *Confirm that* $A\vec{x} = 6\vec{x}$ *for every vector in* $\mathrm{Span}\{\vec{v}\}$, *where* $\vec{v}$ *is the vector you found in part 1. above.*

3. *Compute the matrix* $A - (-3)I_2$.

4. *Find a basis for $\mathcal{N}(A-(-3)I_2)$, i.e., the null space of the matrix that you computed in part 3. above.*

5. *Show that if $\vec{x}$ is in $\mathcal{N}(A-(-3)I_2)$, then $A\vec{x} = -3\vec{x}$. (Hint: start by taking $\vec{x}$ to be the basis element you found in part 4. above.)*

Exercise 6.0.3. *Diagonal matrices are particularly easy to work with. Consider the 3×3 diagonal matrix $A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$ with a, b, and c some real numbers. Show that there are three vectors, say $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_3$, such that*

$$A\vec{v}_1 = a\vec{v}_1, \quad A\vec{v}_2 = b\vec{v}_2, \quad \text{and} \quad A\vec{v}_3 = c\vec{v}_3.$$

6.1 The Determinant

A determinant is a function that assigns a scalar value to a square matrix—i.e., the determinant function takes a square matrix as its input and produces a real number as its output. While the determinant function can be associated with various geometric and algebraic considerations, one of its most useful properties is its association with invertibility. Specifically, the determinant of a square matrix will be zero if and only if the matrix is not invertible[1]. While the general formulation for the determinant of an $n\times n$ matrix is not particularly intuitive, we can arrive at the determinant of a 2×2 matrix by focusing on the question of invertibility.

Definition 6.1.1. *Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. The* **determinant** *of A, denoted $\det(A)$, is the number*

$$\det(A) = ad - bc.$$

Example 6.1.1. *Evaluate the determinant of each of the matrices*

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

[1] We rarely have need to consider 1×1 matrices which can be associated with scalars for practical purposes. If $A = [a_{11}]$ is a 1×1 matrix, we will define its determinant to be the value a_{11}.

2. $B = \begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix}$

3. I_2.

We can apply our formula from Definition 6.1.1.

1. $\det(A) = 1(5) - (-3)(2) = 11$

2. $\det(B) = 1(9) - (-3)(3) = 0$

3. $\det(I_2) = \det\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = 1(1) - 0(0) = 1$

Exercise 6.1.1. *Evaluate the determinant of each of the matrices*

1. $A = \begin{bmatrix} 2 & -4 \\ 6 & 10 \end{bmatrix}$

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

3. $C = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ *where θ is a real number.*

Exercise 6.1.2.

1. *Show that for* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $\det(2A) = 4\det(A)$.

2. *Show that for* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $\det(3A) = 9\det(A)$.

3. *Can you make a conjecture about the relationship between* $\det(kA)$ *and* $\det(A)$ *for a* 2×2 *matrix* A *and a scalar* k*?*

Remark 6.1.1. *Another common notation used to denote a determinant is a pair of vertical bars (resembling absolute value bars),* $\det(A) = |A|$*. This makes it important to write delimiters clearly, especially when writing by hand. The object* $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *is a* 2×2 *matrix, whereas the object* $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ *is a scalar.*

While the relationship between the invertibility of our matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and the number $\det(A) = ad - bc$, may not be immediately apparent, we can deduce the relationship by appealing to the fact that A is invertible if and only if $\text{rref}(A) = I_2$, and this requires both columns of A to be pivot columns. In particular, if $a = 0$ and $c = 0$, then the first column of A is not a pivot column, and A is not invertible. If $a = c = 0$, then the value $ad - bc = 0$. Now, invertibility requires at least one of a or c to be nonzero. Let's assume that $a \neq 0$ and set up the multiply augmented matrix $\left[A \mid I_2\right]$. If we perform a few operations with the goal of reducing A to its rref,

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right] \qquad aR_2 \to R_2 \qquad \left[\begin{array}{cc|cc} a & b & 1 & 0 \\ ac & ad & 0 & a \end{array}\right]$$

$$-cR_1 + R_2 \to R_2 \quad \left[\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & ad - bc & -c & a \end{array}\right].$$

the number $ad - bc$ appears in the second column. The second column will be a pivot column (hence A will be invertible) if and only if this number is nonzero. For the matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we have found that

- if $a = c = 0$, then $ad - bc = 0$, and A is not invertible;
- if $a \neq 0$, then A is invertible if and only if $ad - bc \neq 0$.

Exercise 6.1.3. *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$*, and suppose* $\det(A) \neq 0$*. Show that*

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Note that this provides a quick formula[2] *for the inverse of a* 2×2 *matrix.*

A common approach to computing the determinant for square matrices with 3 or more columns is based on recursion. The determinant of a 3×3 matrix is computed as the weighted sum of determinants of 2×2 matrices. The determinant of a 4×4 matrix is computed as the weighted sum of determinants of 3×3 matrices, and so forth. This approach is called a *Laplace expansion*; it is also referred to as a *cofactor expansion*.

Let us consider a 3×3 matrix,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Let A_{ij} denote the 2×2 matrix obtained from A by eliminating the i^{th} row and the j^{th} column. For example, the matrix A_{23} would be obtained from A by eliminating the second row and the third column

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \to \quad A_{23} = \begin{bmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{bmatrix}.$$

Similarly, the matrix A_{31} is the 2×2 matrix obtained from A by eliminating the third row and first column

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \to \quad A_{31} = \begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix}$$

Since each A_{ij} is a 2×2 matrix, we can use Defintion 6.1.1 to evaluate its determinant $\det(A_{ij})$. We have the following definition of the determinant of a 3×3 matrix.

Definition 6.1.2. *Let* $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$. *The* **determinant** *of* A*,*

[2] There is an analogous formulation for the inverses of larger matrices, but it is computationally intensive. For example, the corresponding formula for the inverse of a 3×3 requires computation of nine 2×2 determinants plus the determinant of the 3×3 matrix. Our row reduction procedure on a multiply augmented matrix is still the practical choice.

denoted $\det(A)$, *is the number*

$$\begin{aligned}\det(A) &= a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13}) \\ &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33}\end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33}\end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32}\end{vmatrix}.\end{aligned}$$

Example 6.1.2. *Compute the determinant of the matrix* $A = \begin{bmatrix} 1 & 3 & -2 \\ 4 & 0 & 1 \\ 2 & 2 & 5 \end{bmatrix}$.

We can simply apply the formula in Definition 6.1.2. Note that

$$A_{11} = \begin{bmatrix} 0 & 1 \\ 2 & 5 \end{bmatrix}, \quad so \quad \det(A_{11}) = 0(5) - 2(1) = -2.$$

$$A_{12} = \begin{bmatrix} 4 & 1 \\ 2 & 5 \end{bmatrix}, \quad so \quad \det(A_{12}) = 4(5) - 2(1) = 18.$$

$$A_{13} = \begin{bmatrix} 4 & 0 \\ 2 & 2 \end{bmatrix}, \quad so \quad \det(A_{13}) = 4(2) - 2(0) = 8.$$

Then

$$\begin{aligned}\det(A) &= a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13}) \\ &= 1(-2) - 3(18) + (-2)(8) \\ &= -72.\end{aligned}$$

Exercise 6.1.4. *Evaluate the determinant of each* 3×3 *matrix.*

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

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

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

Exercise 6.1.5. *Suppose* $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$. *Show that* $\det(A) = a_{11}a_{22}a_{33}$.

6.1.1 Cofactors and the Determinant of an $n \times n$ Matrix

Now that we have a formulation for the determinant of a 3×3 matrix, we can extend this to obtain a determinant formula for larger matrices. For an $n \times n$ matrix, A, we will continue to use the notation A_{ij} to denote the $(n-1) \times (n-1)$ matrix obtained from A by removing the i^{th} row and the j^{th} column. There is a special name for the determinant of such a submatrix A_{ij}; it is called a **minor**.

Definition 6.1.3. *Let A be an $n \times n$ matrix, $n \geq 2$. The $ijth$* **minor** *of A is the determinant of the $(n-1) \times (n-1)$ matrix A_{ij}. That is, $\det(A_{ij})$ is the $ijth$ minor of A.*

For each of the n^2 entries, a_{ij}, in a matrix A, we have a corresponding minor $\det(A_{ij})$. In the determinant formula of Definition 6.1.2, these minors appear with a factor of either 1 or -1. A minor with the appropriate factor is called a cofactor.

Definition 6.1.4. *Let A be an $n \times n$ matrix, $n \geq 2$.*

$$ijth \textbf{ cofactor } of\ A \;=\; (-1)^{i+j}\det(A_{ij}).$$

Note that the factor of $+1$ or -1 is determined by the position of a_{ij} in the matrix, and these follow a predictable, alternating pattern starting with $+1$ in the top left corner. That is, we can asign the correct sign according to the pattern seen here:

$$\begin{bmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Having defined cofactors, we see that the determinant of a 3×3 matrix in Definition 6.1.2 obtained by multiplying each entry in the first row with its corresponding cofactor and adding the results. We now define the determinant of an $n \times n$ matrix.

Definition 6.1.5. *Let $A = [a_{ij}]$ be an $n \times n$ matrix. The determinant of A, denoted $\det(A)$ is given by*

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j}). \tag{6.2}$$

The sum in equation (6.2) is called a **cofactor expansion** *across the first row of A.*

Example 6.1.3. *Evaluate the determinant of* $A = \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 3 & 2 & 2 \\ 1 & -1 & 0 & 4 \\ -1 & 2 & 2 & 1 \end{bmatrix}$.

Solution: *The cofactor expansion will require the determinants of the following four* 3×3 *matrices:*

$$A_{11} = \begin{bmatrix} 3 & 2 & 2 \\ -1 & 0 & 4 \\ 2 & 2 & 1 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 0 & 2 & 2 \\ 1 & 0 & 4 \\ -1 & 2 & 1 \end{bmatrix},$$

$$A_{13} = \begin{bmatrix} 0 & 3 & 2 \\ 1 & -1 & 4 \\ -1 & 2 & 1 \end{bmatrix}, \text{ and } A_{14} = \begin{bmatrix} 0 & 3 & 2 \\ 1 & -1 & 0 \\ -1 & 2 & 2 \end{bmatrix}.$$

Applying the cofactor expansion for each, we obtain

$$\det(A_{11}) = -10, \quad \det(A_{12}) = -6, \quad \det(A_{13}) = -13, \quad \text{and} \quad \det(A_{14}) = -4.$$

Then

$$\begin{aligned} \det(A) &= a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13}) - a_{14}\det(A_{14}) \\ &= (1)(1)(-10) + (-1)(2)(-6) + (1)(-1)(-13) + (-1)(0)(-4) \\ &= 15 \end{aligned}$$

6.1.2 Some Properties of Determinants

The determinant of a 4×4 matrix requires computation of the determinants of four 3×3 matrices, each of which requires computing the determinant of three 2×2 matrices. Despite the existence of a few applications, the computational expense associated with taking a determinant limits its use for almost all but the smallest of matrices. Nevertheless, the determinant will provide us with a useful tool for our main purpose of this chapter, and that is finding eigenvalues and eigenvectors and performing certain matrix decompositions. A number of properties of determinants can be used to our advantage, including some that may simplify taking a determinant.

The formulation of the determinant of an $n \times n$ matrix given in Definition 6.1.5 is stated in terms of a cofactor expansion across the first row of the matrix. However, the determinant can be evaluated by a cofactor expansion across any row or down any column of the matrix.

Property 6.1. *For $n \times n$ matrix A, we can compute $\det(A)$ using a cofactor expansion across the i^{th} row*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}). \tag{6.3}$$

Similarly, we can compute $\det(A)$ using a cofactor expansion down the j^{th} column

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}). \tag{6.4}$$

Note that in the summation in equation (6.3), the value of i is fixed so that the cofactor expansion is computed by multiplying each entry in the i^{th} row by its cofactor and summing. In the summation in equation (6.4), the value of j is fixed. In this formulation, the expansion is computed by multiplying each entry in the j^{th} column by its cofactor and summing.

Example 6.1.4. *In Example 6.1.3, the determinant of the matrix*

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

was computed by cofactor expansion across the first row. Let's compute the determinant by a cofactor expansion down the third column. This means that we will fix the j value as 3 and compute

$$\det(A) = (-1)^{1+3} a_{13} \det(A_{13}) + (-1)^{2+3} a_{23} \det(A_{23}) +$$

$$+ (-1)^{3+3} a_{33} \det(A_{33}) + (-1)^{4+3} a_{43} \det(A_{43}).$$

The 3×3 submatrices for this computation are

$$A_{13} = \begin{bmatrix} 0 & 3 & 2 \\ 1 & -1 & 4 \\ -1 & 2 & 1 \end{bmatrix}, \quad A_{23} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & -1 & 4 \\ -1 & 2 & 1 \end{bmatrix}$$

$$A_{33} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 2 \\ -1 & 2 & 1 \end{bmatrix}, \quad \textit{and} \quad A_{43} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 2 \\ 1 & -1 & 4 \end{bmatrix},$$

with determinants

$$\det(A_{13}) = -13, \quad \det(A_{23}) = -19, \quad \det(A_{33}) = -5, \quad \textit{and} \quad \det(A_{43}) = 18.$$

This gives

$$\det(A) = (1)(-1)(-13) + (-1)(2)(-19) + (1)(0)(-5) + (-1)(2)(18) = 15.$$

Note that the intermediate computations are different, but the value of $\det(A)$ *matches that found in Example 6.1.3 (as it must!).*

Exercise 6.1.6. *Find the determinant of the matrix* $A = \begin{bmatrix} 1 & 2 & -1 \\ 3 & 4 & 0 \\ 2 & -2 & 3 \end{bmatrix}$ *by computing a cofactor expansion*

1. *across the second row,*
2. *down the first column,*
3. *across the third row.*

Since we can choose to use a cofactor expansion across any row or down any column, we can take advantage of the presence of zeros. There's no need to compute a cofactor if we will multiply it by zero, so we may be able to minimize the amount of work.

Exercise 6.1.7. *Find the determinant of each matrix using a cofactor expansion that minimizes the computations.*

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

Exercise 6.1.8. *Suppose A is an $n \times n$ matrix, and A has a row or a column vector of all zeros. Explain why* $\det(A) = 0$.

There are a couple of immediate consequences of Property 6.1. One of these is suggested in Exercise 6.1.8, and the other follows from the relationship between the rows and columns of a matrix A and its transpose A^T.

Property 6.2. *Let A be an $n \times n$ matrix.*

- *If $\vec{0}_n$ is a row vector or a column vector of A, then* $\det(A) = 0$.
- $\det(A^T) = \det(A)$.

Since we can choose to take the determinant using a cofactor expansion across any row or down any column, the structure of a matrix can provide a simpler formulation. Triangular matrices are an example of a special structure. We say that the matrix $A = [a_{ij}]$ is **upper triangular** if $a_{ij} = 0$ for all $i > j$. As the name suggests, an upper triangular matrix has all of its nonzero entries in the upper right triangular area of the matrix. Similarly, the matrix $A = [a_{ij}]$ is called **lower triangular** if $a_{ij} = 0$ for all $i < j$. A lower triangular matrix is readily identified by the presence of all of its nonzero entries in the lower left triangular area of the matrix. A matrix that is both upper triangular and lower triangular is called a **diagonal** matrix.

$$\underset{\text{upper triangular}}{\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}} \quad \underset{\text{lower triangular}}{\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}} \quad \underset{\text{diagonal}}{\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}}$$

To take the determinant of a triangular matrix, we can choose a row or column with only one nonzero entry at each step in the iterative process. This leads to a very simple formula for the determinant of such a matrix.

Property 6.3. *If $A = [a_{ij}]$ is a triangular matrix (upper, lower or diagonal), then the determinant of A is the product of its diagonal entries,*

$$\det(A) = a_{11}a_{22}\cdots a_{nn}.$$

You may recognize the format of an upper triangular matrix from our work with row echelon forms. Unfortunately, it is generally not the case that a matrix A and an echelon form such as $\text{rref}(A)$ have the same determinant. However, we do know how each of the three elementary row operations affects the determinant, and this provides a process by which we can deduce the determinant of a matrix A by considering the determinant of a row equivalent matrix having an advantageous structure.

Property 6.4. *Suppose A is an $n \times n$ matrix.*

- *If B is obtained from A by performing one row scaling, $kR_i \to R_i$, then* $\det(B) = k\det(A)$.
- *If B is obtained from A by performing one row swap, $R_i \leftrightarrow R_j$, then* $\det(B) = -\det(A)$.
- *If B is obtained from A by performing one row replacement, $kR_i+R_j \to R_j$, then* $\det(B) = \det(A)$.

The power of Property 6.4 is that it allows us to use Gaussian elimination, a less computationally expensive process, to reduce a matrix to a row equivalent echelon form (i.e., an upper triangular matrix). This requires that the operations are recorded so that the determinant of the original matrix can be deduced from the determinant of the resulting echelon matrix.

Exercise 6.1.9. *Confirm each of the three statements in Property 6.4 for a 2×2 matrix* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

Exercise 6.1.10. *Suppose A is a 4×4 matrix that is row equivalent to the matrix*

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

If the following row operations were performed on A to produce B, determine $\det(A)$.

- $-2R_1 + R_2 \to R_2$
- $R_3 \leftrightarrow R_4$
- $3R_2 + R_3 \to R_3$
- $\frac{1}{2}R_3 \to R_3$
- $-R_2 + R_4 \to R_4$

Exercise 6.1.11. *If A is an $n \times n$ matrix, explain why* $\det(kA) = k^n \det(A)$ *for scalar k.*

There is a rather surprising property of the determinant that we will find useful for our study of eigenvalues and associated matrix decompositions. The determinant of a product, AB, of a pair of matrices is equal to the product of their determinants.

Property 6.5. *If A and B are $n \times n$ matrices, then*

$$\det(AB) = \det(A)\det(B).$$

Example 6.1.5. *A proof of Property 6.5 is often done by induction on the size of the matrix along with the use of simple matrices known as* elementary *matrices (these are matrices obtained by performing one row operation on the identity I_n). For 2×2 matrices, we can establish this property by direct computation. Let*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} e & f \\ g & h \end{bmatrix}.$$

Then $\det(A) = ad - bc$, $\det(B) = eh - gf$, *and*

$$AB = \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix}.$$

Then note that

$$\begin{aligned} \det(AB) &= (ae + bg)(cf + dh) - (ce + dg)(af + bh) \\ &= aecf + bgcf + aedh + bgdh - ceaf - dgaf - cebh - dgbh \\ &= bgcf + aedh - dgaf - cebh \\ &= ad(eh - gf) - bc(eh - gf) \\ &= (ad - bc)(eh - gf) \\ &= \det(A)\det(B). \end{aligned} \tag{6.5}$$

Exercise 6.1.12. *For each pair of matrices A and B, evaluate the products AB and BA. Compute the determinants* $\det(A)$, $\det(B)$, $\det(AB)$, *and* $\det(BA)$ *and confirm that* $\det(AB) = \det(A)\det(B) = \det(BA)$.

1. $A = \begin{bmatrix} 1 & 2 \\ -4 & 3 \end{bmatrix}$ *and* $B = \begin{bmatrix} 3 & 1 \\ 2 & 5 \end{bmatrix}$.

2. $A = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & 1 \\ 2 & -2 & 4 \end{bmatrix}$ *and* $B = \begin{bmatrix} 6 & 0 & -1 \\ 2 & 2 & 2 \\ -1 & 0 & 3 \end{bmatrix}$.

The most critical property of the determinant for our present purpose is its relationship to the invertibility of a matrix.

Theorem 6.1.1. *Let A be an $n \times n$ matrix. A is invertible if and only if* $\det(A) \neq 0$.

Proof. We recall from Theorem 3.9.3 that an $n \times n$ matrix A is invertible if and only if $\mathrm{rref}(A) = I_n$. The row reduction process consists of some sequence of the three elementary row operations (scaling, swapping, replacement), and from Property 6.4, each such operation scales the determinant by a nonzero scalar. If $\mathrm{rref}(A)$ is obtained from A by performing a sequence of p elementary row operations, then

$$\det(\mathrm{rref}(A)) = \alpha_1 \alpha_2 \cdots \alpha_p \det(A),$$

where each factor $\alpha_i \neq 0$ (each factor is either 1, -1, or some nonzero scaling factor). Hence

$$\det(A) = \hat{\alpha} \det(\mathrm{rref}(A)), \quad \text{where} \quad \hat{\alpha} \neq 0.$$

If A is invertible, then

$$\det(A) = \hat{\alpha} \det(\mathrm{rref}(A)) = \hat{\alpha} \det(I_n) = \hat{\alpha} \neq 0.$$

If A is not invertible, then $\mathrm{rref}(A) \neq I_n$, and $\mathrm{rref}(A)$ has at least one row of all zero so that $\det(\mathrm{rref}(A)) = 0$. In this case,

$$\det(A) = \hat{\alpha} \det(\mathrm{rref}(A)) = \hat{\alpha}(0) = 0.$$

□

Example 6.1.6. *Suppose* $A = \begin{bmatrix} 2-\lambda & 3 \\ 1 & -1-\lambda \end{bmatrix}$, *where λ is a real number. Determine all values of λ, if any, such that A is not invertible.*

We can use the fact that A is not invertible if its determinant is zero. We obtain an equation that we can solve for λ.

$$\det(A) = (2-\lambda)(-1-\lambda) - 1(3) = \lambda^2 - \lambda - 5.$$

Setting $\det(A) = 0$*, we get a quadratic equation*

$$\lambda^2 - \lambda - 5 = 0,$$

with two solutions

$$\lambda = \frac{1+\sqrt{21}}{2}, \quad \textit{or} \quad \lambda = \frac{1-\sqrt{21}}{2}.$$

These are the only two values of λ *for which the matrix* A *is not invertible.*

Exercise 6.1.13. *For each matrix* A*, determine all values of* λ*, if any, such that* A *is not invertible.*

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

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

3. $A = \begin{bmatrix} 3-\lambda & 0 \\ 2 & 3-\lambda \end{bmatrix}$

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

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

6.2 Eigenvalues & Eigenvectors

In Chapter 5, we learned that one of the defining features of a linear transformation is that it maps a line to a line or to a point. For the various examples of maps from $R^2 \to R^2$, we can even plot lines and their images to better understand the action of a given linear transformation. As a general rule, we don't expect a linear transformation to map a given line back to itself, but it may happen, and we might consider such an action as a characteristic of the transformation. We opened this chapter with an example of a linear transformation, $T(\vec{x}) = \begin{bmatrix} 5 & -1 \\ 3 & 1 \end{bmatrix} \vec{x}$, and the observation that there was a special set of vectors, namely Span$\{\langle 1,3\rangle\}$, for which the matrix-vector product, $A\vec{x}$, is equivalent to the scalar multiplication of the vector $2\vec{x}$. In Exercise 6.0.1, you might have deduced that there is another collection of vectors, namely Span$\{\langle 1,1\rangle\}$, for which this transformation scales, but does not change direction. The scalar for those vectors is 4. For this particular transformation, the vectors in Span$\{\langle 1,3\rangle\}$ with special scaling factor 2 and the vectors in Span$\{\langle 1,1\rangle\}$ with special scaling factor 4 are the only vectors in R^2 with this special property. The prefix *Eigen*, from the German for *own* or *characteristic*, is used to describe these special vectors and scaling factors.

Definition 6.2.1. *Let A be an $n\times n$ matrix. An* **eigenvalue** *of A is a scalar λ for which there exists a nonzero vector $\vec{x}$ such that*

$$A\vec{x} = \lambda\vec{x}. \tag{6.6}$$

For a given eigenvalue λ, a nonzero vector $\vec{x}$ satisfying equation (6.6) is called an **eigenvector** *corresponding to the eigenvalue λ.*

Remark 6.2.1. *A perhaps subtle but critical feature of Definition 6.2.1 is that eigenvectors are nonzero vectors. The equation $A\vec{x} = \lambda\vec{x}$ is trivially satisfied by $\vec{x} = \vec{0}_n$ no matter what the value of the scalar λ, but the zero vector is not an eigenvector. We place no such restriction on eigenvalues. That is, an eigenvalue λ can be any real number, including zero. (For certain applications, we may be interested in allowing λ to be a complex number, say $\lambda = a + ib$ where a and b are real numbers and $i^2 = -1$.)*

Example 6.2.1. *Let* $A = \begin{bmatrix} -4 & 1 \\ 7 & 2 \end{bmatrix}$.

1. *Show that $\lambda = -5$ is an eigenvalue of A by finding a nonzero vector $\vec{x}$ such that $A\vec{x} = -5\vec{x}$.*

2. *Show that $\vec{x} = \langle 1,7\rangle$ is an eigenvector of A by showing that there is a scalar λ such that $A\langle 1,7\rangle = \lambda\langle 1,7\rangle$.*

For part 1., we can obtain a system of equations. Let $\vec{x} = \langle x_1, x_2\rangle$, and suppose $A\vec{x} = -5\vec{x}$. We have

$$A\vec{x} = \langle -4x_1 + x_2, 7x_1 + 2x_2\rangle = -5\langle x_1, x_2\rangle.$$

Equating each of the entries in these vectors produces a system of equations

$$\begin{array}{rcrcr} -4x_1 & + & x_2 & = & -5x_1 \\ 7x_1 & + & 2x_2 & = & -5x_2 \end{array}.$$

Writing this in the more traditional format with x_1 and x_2 on the left side, we see that this is actually a homogeneous system

$$\begin{array}{rcrcr} x_1 & + & x_2 & = & 0 \\ 7x_1 & + & 7x_2 & = & 0 \end{array}. \tag{6.7}$$

It is worth noting at this step that we have a homogeneous system with a coefficient matrix that is not our original matrix A. The coefficient matrix is the matrix obtained from A by subtracting -5 from the diagonal entries,

$$\begin{bmatrix} 1 & 1 \\ 7 & 7 \end{bmatrix} = \begin{bmatrix} -4-(-5) & 1 \\ 7 & 2-(-5) \end{bmatrix} = \begin{bmatrix} -4 & 1 \\ 7 & 2 \end{bmatrix} - \begin{bmatrix} -5 & 0 \\ 0 & -5 \end{bmatrix} = A-(-5)I_2.$$

We are only interested in nontrivial solutions to the system 6.7, and this will only be possible if our coefficient matrix, $A-(-5)I_2$, is not invertible. we can solve system 6.7 using an augmented matrix and row reduction in the traditional way.

$$\left[\begin{array}{cc|c} 1 & 1 & 0 \\ 7 & 7 & 0 \end{array}\right] \xrightarrow{rref} \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

The coefficient matrix has only one pivot column (so it is not invertible), and solutions to system 6.7 are vectors of the form $\vec{x} = t\langle -1, 1\rangle$. We can take any nonzero value[3] *of t to obtain an eigenvector of A corresponding to the eigenvalue $\lambda = -5$. A simple examples is $\vec{x} = \langle -1, 1\rangle$. Before moving on to the second part of the example, let's verify that our solution satisfies $A\vec{x} = -5\vec{x}$. Note that*

$$A\vec{x} = \langle -4(-1)+1(1), 7(-1)+2(1)\rangle = \langle 5, -5\rangle = -5\langle -1, 1\rangle = -5\vec{x},$$

as expected.

For part 2., we can perform the product $A\langle 1, 7\rangle$.

$$A\langle 1, 7\rangle = \langle -4(1)+1(7), 7(1)+2(7)\rangle = \langle 3, 21\rangle = 3\langle 1, 7\rangle.$$

We see that for $\vec{x} = \langle 1, 7\rangle$, $A\vec{x} = \lambda\vec{x}$ where the value $\lambda = 3$.

[3] Taking $t = 0$ does produce a solution to the system of equations 6.7. However, this choice produces the trivial solution which, by definition, is not an eigenvector.

Exercise 6.2.1. *Let* $A = \begin{bmatrix} 3 & -1 \\ 5 & -3 \end{bmatrix}$.

1. *Show that* $\lambda = 2$ *is an eigenvalue of* A *by finding a nonzero vector* $\vec{x}$ *such that* $A\vec{x} = 2\vec{x}$.

2. *Show that* $\vec{x} = \langle 1, 5 \rangle$ *is an eigenvector of* A *by finding a scalar* λ *such that* $A\vec{x} = \lambda\vec{x}$.

3. *Show that the number* $\lambda = 3$ *is not an eigenvalue of* A. *(Hint: Show that* $A\vec{x} = 3\vec{x}$ *has no nontrivial solutions.)*

Exercise 6.2.2. *We've seen that if* $(\lambda, \vec{x})$ *is an eigenvalue-eigenvector pair for a matrix* A*, then* $A\vec{x}$ *is in* $\mathrm{Span}\{\vec{x}\}$. *Consider the transformation* $R_{90^\circ}(\vec{x}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \vec{x}$ *that rotates a vector in* R^2 *by* 90° *counterclockwise. Explain why there are no (real) numbers* λ *that are eigenvalue of the matrix* $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

6.2.1 The Characteristic Equation

As the examples and exercises suggest, given a possible eigenvalue, we can use existing tools to find corresponding eigenvectors (and vice versa). We still require some process by which to determine whether a given matrix has any eigenvalues and if so, to determine what they are. For $n \times n$ matrix A, we are interested in the equation

$$A\vec{x} = \lambda\vec{x}. \tag{6.8}$$

Here, we are equating the vector in R^n resulting from the matrix-vector product $A\vec{x}$ to the vector in R^n resulting from scaling the vector $\vec{x}$. Equation (6.8) is different from matrix-vector equations we've previously encountered. Our unknown vector $\vec{x}$ appears on both sides of the equation along with an unknown scalar. Moreover, the two sides of equation (6.8) involve different types of products. To use our existing tools to manipulate this equation, it is advantageous to rephrase the right side of this equation as a matrix-vector product. We need a matrix that will scale every vector by the (as yet unknown) scalar λ. The matrix λI_n does precisely that, so we can write equation (6.8) as

$$A\vec{x} = \lambda I_n \vec{x},$$

and rearrange to obtain the homogeneous equation

$$A\vec{x} - \lambda I_n\vec{x} = \vec{0}_n.$$

Factoring the vector $\vec{x}$, we seek scalar(s) λ such that the homogeneous equation

$$(A - \lambda I_n)\vec{x} = \vec{0}_n \tag{6.9}$$

has nontrivial solutions. At this point, an option is to construct the augmented matrix $\left[A - \lambda I_n \mid \vec{0}_n\right]$ and commence with row reduction. If A is a small matrix (say 2×2), and we exercise patience and caution, this is a legitimate (if not somewhat unattractive) approach. We will take another approach. A direct consequence of Theorem 3.9.3 is that the equation $(A - \lambda I_n)\vec{x} = \vec{0}_n$ has nontrivial solutions if and only if the matrix $A - \lambda I_n$ is not invertible. By Theorem 6.1.1, we know that $A - \lambda I_n$ is not invertible if and only if its determinant is zero. Since the determinant is scalar valued, this provides us with a scalar valued equation for the eigenvalues, λ.

In Exercise 6.1.13, you had the opportunity to take the determinant of a few matrices that had "some number minus λ" in each of the diagonal entries. You probably noted that this always resulted in a polynomial. The degree of the resulting polynomial matched the number of diagonal entries, which of course coincides with the size of the square matrix. Given the cofactor expansion formulation of the determinant, it will necessarily be that $\det(A - \lambda I_n)$ will be an n^{th} degree polynomial in the variable λ.

Definition 6.2.2. *Let A be an $n \times n$ matrix. The function*

$$P_A(\lambda) = \det(A - \lambda I_n)$$

is called the **characteristic polynomial** *of the matrix A. The equation*

$$P_A(\lambda) = 0, \quad i.e., \quad \det(A - \lambda I_n) = 0$$

is called the **characteristic equation** *of the matrix A.*

As the name suggests, the characteristic polynomial of a matrix is a polynomial. Its degree, as you've seen through examples (and as can be proven by induction on the size of the matrix), is equal to the size of the matrix. The characteristic polynomial provides a tool that we can use to determine the eigenvalues of a matrix.

Theorem 6.2.1. *Let A be an $n \times n$ matrix, and let $P_A(\lambda)$ be the characteristic polynomial of A. The number λ_0 is an eigenvalue of A if and only if $P_A(\lambda_0) = 0$. That is, λ_0 is an eigenvalue of A if and only if it is a root of the characteristic equation* $\det(A - \lambda I_n) = 0$.

Proof. Suppose λ_0 is an eigenvalue of the matrix A. Then there exists a nonzero vector $\vec{x}$ such that

$$A\vec{x} = \lambda_0\vec{x}.$$

Hence $\vec{x}$ is a nonzero solution of the homogeneous matrix-vector equation

$$(A - \lambda_0 I_n)\vec{x} = \vec{0}_n.$$

The existence of a nontrivial solution to this homogeneous equation implies that the matrix $A - \lambda_0 I_n$ is not invertible. By Theorem 6.1.1, this matrix has determinant zero. That is,

$$\det(A - \lambda_0 I_n) = 0, \quad \text{i.e.,} \quad P_A(\lambda_0) = 0.$$

Conversely, suppose $P_A(\lambda_0) = 0$ for some number λ_0. Since the determinant of $A - \lambda_0 I_n$ is zero, $A - \lambda_0 I_n$ is not invertible. Hence there exists a nonzero vector $\vec{x}$ such that

$$(A - \lambda_0 I_n)\vec{x} = \vec{0}_n.$$

We can rearrange this equation to find that $\vec{x}$ is nonzero vector such that

$$A\vec{x} = \lambda_0\vec{x},$$

and conclude that λ_0 is an eigenvalue of A. □

Given an $n \times n$ matrix A, Theorem 6.2.1 is applied to arrive at an equation for the eigenvalues. Once any eigenvalue λ_0 is identified, associated eigenvectors are obtained by characterizing the null space of the matrix $A - \lambda_0 I_n$.

Example 6.2.2. *Identify all eigenvalues of the matrix A and find an associated eigenvector for each eigenvalue.*

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

First, we find the characteristic polynomial for A. Note that

$$A - \lambda I_2 = \begin{bmatrix} -2 & 3 \\ 6 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2-\lambda & 3 \\ 6 & 1-\lambda \end{bmatrix}.$$

Taking the determinant,

$$\det(A - \lambda I_2) = (-2-\lambda)(1-\lambda) - 6(3) = \lambda^2 + \lambda - 20.$$

The quadratic factors readily as $P_A(\lambda) = (\lambda-4)(\lambda+5)$, *and we have two solutions to the characteristic equation* $(\lambda-4)(\lambda+5) = 0$. *We can label these*

$$\lambda_1 = 4, \quad \textit{and} \quad \lambda_2 = -5.$$

Next, for each eigenvalue, we set up the equation $(A - \lambda_i I_2)\vec{x} = \vec{0}_2$ *and identify solutions. For* $\lambda_1 = 4$, *the coefficient matrix*

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

Setting up the augmented matrix and performing the row reduction,

$$\left[\begin{array}{rr|r} -6 & 3 & 0 \\ 6 & -3 & 0 \end{array}\right] \quad \xrightarrow{rref} \quad \left[\begin{array}{rr|r} 1 & -1/2 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

The solutions, $\vec{x} = \langle x_1, x_2 \rangle$ *are of the form* $\vec{x} = t\left\langle \frac{1}{2}, 1 \right\rangle$, $t \in R$. *Taking* $t = 2$ *gives a representative eigenvector* $\vec{x}_1 = \langle 1, 2 \rangle$. *Repeating the procedure for* $\lambda_2 = -5$, *we find that solutions of the homogeneous equation* $(A - (-5)I_2)\vec{x} = \vec{0}$ *are of the form* $\vec{x} = s\langle -1, 1 \rangle$. *A representative eigenvector (selecting* $s = 1$*) is* $\vec{x}_2 = \langle -1, 1 \rangle$. *To summarize, we found the eigenvalue-eigenvector pairs,*

$$\lambda_1 = 4, \quad \vec{x}_1 = \langle 1, 2 \rangle, \quad \textit{and} \quad \lambda_2 = -5, \quad \vec{x}_2 = \langle -1, 1 \rangle.$$

Exercise 6.2.3. *For each matrix, determine all eigenvalues and for each eigenvalue, find a corresponding eigenvector.*

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

2. $A = \begin{bmatrix} 3 & 0 \\ 2 & 3 \end{bmatrix}$

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

Determining the eigenvalues of an $n \times n$ matrix A requires us to solve an n^{th} degree polynomial equation $P_A(\lambda) = 0$. This is a straightforward task when $n = 2$ (we can always apply the quadratic formula), but can be quite the challenge—perhaps impossible without the help of computational software—for large n. The eigenvalues of select matrices, specifically triangular matrices, are readily identified as is stated in the following Theorem (the proof of which is left as an exercise).

Theorem 6.2.2. *If $A = [a_{ij}]$ is an $n \times n$ triangular matrix (upper, lower, or diagonal), the eigenvalues of A are its diagonal entries. That is, the eigenvalues, $\lambda_i = a_{ii}$ for $i = 1, \ldots, n$.*

Before we proceed, we state one additional theorem on the connection between the invertibility of a matrix and its eigenvalues.

Theorem 6.2.3. *Let A be an $n \times n$ matrix. Then A is invertible if and only if zero is not an eigenvalue of A.*

Proof. Let A be an $n \times n$ matrix with characteristic polynomial P_A. Suppose A is invertible. Then by Theorem 6.1.1, $\det(A) \neq 0$. Then

$$P_A(0) = \det(A - 0I_n) = \det(A) \neq 0,$$

and $\lambda = 0$ is not a zero of the characteristic polynomial $P_A(\lambda)$ and hence not an eigenvalue of A. Conversely, suppose A is not invertible. Then by Theorem 3.9.3, $\text{rref}(A) \neq I_n$ and there exists a nontrivial solution to the homogeneous equation $A\vec{x} = \vec{0}_n$. Let $\vec{x}_0$ be such a nontrivial solution. Then

$$A\vec{x}_0 = 0\vec{x}_0.$$

That is, zero is an eigenvalue of A. □

6.2.2 Eigenspaces & Eigenbases

The process of identifying eigenvalue-eigenvector pairs starts with finding eigenvalues by solving the characteristic equation. With an eigenvalue in hand, the associated eigenvectors are nontrivial solutions to a specific homogeneous equation. For an $n \times n$ matrix, this tells us that eigenvectors associated with a given eigenvector are all elements of a specific subspace of R^n. We call this subspace an **eigenspace**.

Definition 6.2.3. *Let A be an $n \times n$ matrix and λ_0 be an eigenvalue of A. The* **eigenspace corresponding to the eigenvalue** λ_0 *is the set*

$$E_A(\lambda_0) = \{\vec{x} \in R^n \mid A\vec{x} = \lambda_0 \vec{x}\} = \mathcal{N}(A - \lambda_0 I_n).$$

As a null space, an eigenspace for an $n \times n$ matrix is necessarily a subspace of R^n. The eigenvectors associated with λ_0 are all of the nonzero vectors in this subspace, $E_A(\lambda_0)$. In general, we can characterize a subspace of R^n by a basis, and since $E_A(\lambda_0)$ is the null space of a matrix, we can use our familiar procedure to find a basis.

Example 6.2.3. *It can be shown that the matrix A below has characteristic polynomial $P_A(\lambda) = (2-\lambda)^2(9-\lambda)$, so A has two eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 9$. Find a basis for the eigenspace of A corresponding to the eigenvalue $\lambda_1 = 2$, $E_A(2)$.*

$$A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}.$$

We want to solve the homogeneous equation $(A - 2I_3)\vec{x} = \vec{0}_3$.

$$\left[A - 2I_3 \mid \vec{0}_3\right] = \left[\begin{array}{ccc|c} 4-2 & -1 & 6 & 0 \\ 2 & 1-2 & 6 & 0 \\ 2 & -1 & 8-2 & 0 \end{array}\right] \xrightarrow{rref} \left[\begin{array}{ccc|c} 1 & -1/2 & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

From the rref, we see that $A - 2I_3$ has one pivot column and two non-pivot columns. Any solution, $\vec{x} = \langle x_1, x_2, x_3 \rangle$, of the homogeneous equation $(A - 2I_3)\vec{x} = \vec{0}_3$ will satisfy

$$x_1 = \frac{1}{2}x_2 - 3x_3, \quad \text{with } x_2 \text{ and } x_3 \text{ free.}$$

With two free variables, we can write such an eigenvector as the linear combination of two, linearly independent vectors

$$\vec{x} = s\left\langle \frac{1}{2}, 1, 0 \right\rangle + t\left\langle -3, 0, 1 \right\rangle.$$

A basis for $E_A(2)$ is

$$\left\{\left\langle \frac{1}{2}, 1, 0 \right\rangle, \left\langle -3, 0, 1 \right\rangle\right\}.$$

Example 6.2.4. *The matrix B below has the same characteristic polynomial, $P_B(\lambda) = (2-\lambda)^2(9-\lambda)$, as the matrix in Example 6.2.3.*

$$B = \begin{bmatrix} -12 & 9 & 8 \\ -70 & 32 & 15 \\ 42 & -18 & -7 \end{bmatrix}.$$

Find a basis for the eigenspace, $E_B(2)$, of B corresponding to the eigenvalue $\lambda_1 = 2$.

We proceed as we did in the last example by performing row reduction on the matrix $\left[B - 2I_3|\vec{0}_3\right]$.

$$\left[\begin{array}{ccc|c} -12-2 & 9 & 8 & 0 \\ -70 & 32-2 & 15 & 0 \\ 42 & -18 & -7-2 & 0 \end{array}\right] \xrightarrow{rref} \left[\begin{array}{ccc|c} 1 & 0 & 1/2 & 0 \\ 0 & 1 & 5/3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

We see that there are two basic and one free variable. Solutions $\vec{x} = \langle x_1, x_2, x_3\rangle$ will satisfy

$$x_1 = -\frac{1}{2}x_3, \quad x_2 = -\frac{5}{3}x_3, \quad \text{with } x_3 \text{ free.}$$

A parametric form of the solution is $\vec{x} = t\left\langle -\frac{1}{2}, -\frac{5}{3}, 1\right\rangle$ for $t \in R$. A basis for $E_B(2)$ is

$$\left\{\left\langle -\frac{1}{2}, -\frac{5}{3}, 1\right\rangle\right\}.$$

(Note that if we select $t = -6$, we can take our basis to be the, perhaps more attractive, set $\{\langle 3, 10, -6\rangle\}$.)

The matrices A and B in Examples 6.2.3 and 6.2.4 have the same characteristic polynomial and hence the same eigenvalues. As we saw in these examples, however, the eigenspaces $E_A(2)$ and $E_B(2)$ are not the same. Most notably, these subspaces have different dimensions as evidenced by the different number of basis elements we found. The dimension of an eigenspace is a characteristic of an eigenvalue for a given matrix.

Definition 6.2.4. *Let A be an $n \times n$ matrix and λ_0 be an eigenvalue of A. The dimension of the eigenspace, $\dim(E_A(\lambda_0))$, corresponding to λ_0 is called the* **geometric multiplicity** *of λ_0.*

For the matrix A in Examples 6.2.3, we found a basis for the eigenspace corresponding to $\lambda_1 = 2$ with two vectors making the geometric multiplicity two. The geometric multiplicity of the eigenvalue $\lambda_1 = 2$ for the matrix B in Example 6.2.4 is one; the basis for the corresponding eigenspace contains one vector. The geometric multiplicity of an eigenvalue tells us how many linearly independent eigenvectors correspond to that eigenvalue.

As the root of an n^{th} degree polynomial equation, there is a second type of multiplicity associated with an eigenvalue for a matrix. In general, an n^{th} degree polynomial with real coefficients has at most n real zeros. It may have fewer than n, including non-real complex zeros. This places a limit on the number of eigenvalues that a matrix can have.

Definition 6.2.5. *Let A be an $n \times n$ matrix and λ_0 be an eigenvalue of A. The* **algebraic multiplicity** *of λ_0 is its multiplicity as the root of the characteristic equation $P_A(\lambda) = 0$. That is, if $(\lambda - \lambda_0)^k$ is a factor of $P_A(\lambda)$ and $(\lambda - \lambda_0)^{k+1}$ is not a factor of $P_A(\lambda)$, then the algebraic multiplicity of λ_0 is k.*

The matrices A and B in Examples 6.2.3 and 6.2.4 both have characteristic polynomial $(\lambda - 2)^2(\lambda - 9)$. From this, we see that the eigenvalue $\lambda_1 = 2$ has algebraic multiplicity two, and the eigenvalue $\lambda_2 = 9$ has algebraic multiplicity one. The algebraic multiplicity of an eigenvalue is a limit on the number of linearly independent eigenvectors a matrix may have. In particular, the algebraic multiplicity of an eigenvalue is greater than or equal to its geometric multiplicity.

Exercise 6.2.4. *Consider the pair of matrices*

$$A = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 5 \end{bmatrix}.$$

1. *Find the characteristic polynomials P_A and P_B and show that they are equal, $P_A(\lambda) = P_B(\lambda)$.*

2. *Identify the eigenvalues of A and for each eigenvalue of A determine its algebraic multiplicity and its geometric multiplicity.*

3. *Identify the eigenvalues of B and for each eigenvalue of B determine its algebraic multiplicity and its geometric multiplicity.*

One of the uses of eigenvalues and eigenvectors is that they may allow us to express a linear transformation—i.e., a matrix, using a basis in which the matrix is diagonal. To do this, we need a basis for R^n, and such a basis necessarily contains n linearly independent vectors. Hence the linear dependence or independence of a set of eigenvectors for a matrix is of interest.

Example 6.2.5. *Suppose an $n \times n$ matrix A has distinct eigenvalues λ_1 and λ_2 with corresponding eigenvectors $\vec{x}_1$ and $\vec{x}_2$. Show that the set $\{\vec{x}_1, \vec{x}_2\}$ is linearly independent.*

To show that $\{\vec{x}_1, \vec{x}_2\}$ is linearly independent, let's consider the homogeneous equation

$$c_1\vec{x}_1 + c_2\vec{x}_2 = \vec{0}_n. \tag{6.10}$$

We want to show that the only solution is the trivial one, $c_1 = c_2 = 0$. We can create a system of two new equations. To generate the first, let's multiply both sides of equation (6.10) by the matrix A and make use of the fact that $A\vec{x}_1 = \lambda_1\vec{x}_1$ and $A\vec{x}_2 = \lambda_2\vec{x}_2$. We have

$$A\left(c_1\vec{x}_1 + c_2\vec{x}_2\right) = A\vec{0}_n \quad \Longrightarrow c_1A\vec{x}_1 + c_2A\vec{x}_2 = \vec{0}_n,$$

which gives

$$c_1\lambda_1\vec{x}_1 + c_2\lambda_2\vec{x}_2 = \vec{0}_n. \tag{6.11}$$

Since λ_1 and λ_2 are distinct, at least one of these is nonzero. We can assume that $\lambda_1 \neq 0$. We will create another equation by multiplying equation (6.10) through by λ_1 to obtain

$$c_1\lambda_1\vec{x}_1 + c_2\lambda_1\vec{x}_2 = \vec{0}_n. \tag{6.12}$$

(Note that equations (6.11) and (6.12) differ only in the coefficient of $\vec{x}_2$.) Now, we subtract equation (6.12) from equation (6.11) to obtain

$$c_2\left(\lambda_2 - \lambda_1\right)\vec{x}_2 = \vec{0}_n.$$

Since $\vec{x}_2$ is an eigenvectors, it is not the zero vector. So it must be that $c_2\left(\lambda_2 - \lambda_1\right) = 0$, and since $\lambda_1 \neq \lambda_2$, we see that $c_2 = 0$ necessarily. This means that equation (6.10) is

$$c_1\vec{x}_1 = \vec{0}_n.$$

But as $\vec{x}_1$ is also an eigenvector and necessarily not the zero vector, we have $c_1 = 0$ as well. We have shown that the homogeneous equation (6.10) has only the trivial solution which confirms that the set $\{\vec{x}_1, \vec{x}_2\}$ is linearly independent.

Example 6.11 illustrates that two eigenvectors of a matrix corresponding to different eigenvalues are necessarily linearly independent. More generally, we have the following theorem.

Theorem 6.2.4. *Let $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ be a set of eigenvectors of an $n \times n$ matrix corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Then the set $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ is linearly independent.*

One consequence of Theorem 6.2.4 is that an $n \times n$ matrix with n distinct real eigenvalues is guaranteed to have n linearly independent eigenvectors. If an $n \times n$ matrix has fewer than n distinct real eigenvalues, for example one or more real eigenvalues has algebraic multiplicity two or greater, the matrix may (e.g., the matrix A in Example 6.2.3) or may not (e.g., the matrix B in Example 6.2.4) have n linearly independent eigenvectors. A set of n linearly independent vectors is a basis for R^n. If we can construct a basis for R^n consisting of eigenvectors for a specific matrix, we aptly call such a basis an **eigenbasis**.

Definition 6.2.6. *Let A be an $n \times n$ matrix. If A has n linearly independent eigenvectors, $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ (combined across all eigenvalues), then the set $\mathcal{E}_A = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\}$ is a basis for R^n. The set $\mathcal{E}_A$ is called an* **eigenbasis** *for A.*

Example 6.2.6. *Find an eigenbasis for the matrix A from Example 6.2.3. Recall*

$$A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}.$$

We were given the characteristic polynomial $P_A(\lambda) = (2-\lambda)^2(9-\lambda)$ from which we see that A has two eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 9$. In Example 6.2.3, we found the basis $\left\{\left\langle \frac{1}{2}, 1, 0\right\rangle, \langle -3, 0, 1\rangle\right\}$ for the eigenspace $E_A(2)$. We need to find a basis for the eigenspace $E_A(9)$ corresponding to the other eigenvalue, $\lambda_2 = 9$. We can set up the homogeneous equation $(A - 9I_3)\vec{x} = \vec{0}_3$. We have

$$\left[A - 9I_3 | \vec{0}_3\right] \quad \xrightarrow{rref} \quad \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right],$$

from which we see that solutions have the form $\vec{x} = t\langle 1,1,1\rangle$, $t \in R$. *We can select* $\{\langle 1,1,1\rangle\}$ *as a basis for* $E_A(9)$. *The union of the bases for these two eigenspaces is an eigenbasis for* A.

$$\mathcal{E}_A = \left\{ \left\langle \frac{1}{2}, 1, 0 \right\rangle, \langle -3, 0, 1\rangle, \langle 1, 1, 1\rangle \right\}$$

Whether we can construct an eigenbasis for a particular matrix depends on the geometric multiplicities of its eigenvalues.

Property 6.6. *A matrix* A *has an eigenbasis if and only if the sum of the geometric multiplicities of all of its eigenvalues is* n.

Exercise 6.2.5. *For each of the matrices*

$$A = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 5 \end{bmatrix}.$$

from Exercise 6.2.4, construct an eigenbasis or explain why one does not exist.

6.3 Diagonalization

We've seen that diagonal matrices are particularly easy to work with. The determinant is a simple product, the eigenvalues are the diagonal entries, and as we saw in Section 5.5, computing powers of such a matrix (multiplying it by itself any number of times) doesn't require the numerous operations generally associated with matrix multiplication. Given a matrix A that is not diagonal, we can ask whether there is a diagonal matrix D that is similar to A. Recall from Definition 5.5.1, that a matrix D is said to be similar to A if there exists an invertible matrix C such that $D = C^{-1}AC$.

Definition 6.3.1. *Let* A *be an* $n \times n$ *matrix. We say that* A *is* **diagonalizable** *if there is a diagonal matrix that is similar to* A. *That is,* A *is* **diagonalizable** *if there exists a diagonal matrix* D *and an invertible matrix* C *such that* $D = C^{-1}AC$.

Exercise 6.3.1. *Let* $B = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 5 \end{bmatrix}$ *and* $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}$. *Show that* B *is diagonalizable. To do this, find* C^{-1} *and compute the product* $C^{-1}BC$.

As we will see, similar matrices share properties such as having the same determinant, characteristic equation, and eigenvalues (though they generally have different corresponding eigenvectors).

Theorem 6.3.1. *If A and B are similar $n \times n$ matrices, then* $\det(A) = \det(B)$.

Proof. Let's suppose that A and B are $n \times n$ similar matrices. Then there is an invertible matrix C such that $B = C^{-1}AC$. Hence

$$\det(B) = \det(C^{-1}AC) = \det(C^{-1})\det(A)\det(C),$$

where we used Property 6.5 that says that the determinant of a product is the product of the determinants. The determinant is scalar valued, so the factors on the right side commute and we have

$$\det(B) = \det(C^{-1})\det(C)\det(A).$$

But note that

$$\det(C^{-1})\det(C) = \det(C^{-1}C) = \det(I_n) = 1.$$

Hence

$$\det(B) = 1\det(A) = \det(A),$$

as required. □

A consequence of Theorem 6.3.1 is that in addition to sharing a determinant, similar matrices have the same characteristic equation and hence the same eigenvalues.

Theorem 6.3.2. *If A and B are similar $n \times n$ matrices, then A and B have the same characteristic polynomial and the same eigenvalues with the same algebraic multiplicities and geometric multiplicities.*

Before we prove Theorem 6.3.2, we estabilish the following lemma that tells us that an invertible linear transformation preserves the linear independence of a set of vectors.

Lemma 6.3.1. *Suppose $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ is a linearly independent set of vectors in R^n. If A is an invertible $n \times n$ matrix, then $\{A\vec{x}_1, A\vec{x}_2, \ldots, A\vec{x}_k\}$ is linearly independent.*

Proof. To prove Lemma 6.3.1, suppose the set $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ is linearly independent but that $\{A\vec{x}_1, A\vec{x}_2, \ldots, A\vec{x}_k\}$ is linearly dependent. Then there exists a linear dependence relation

$$c_1 A\vec{x}_1 + c_2 A\vec{x}_2 + \cdots + c_k A\vec{x}_k = \vec{0}_n \tag{6.13}$$

with at least one of the weights $c_i \neq 0$. Since A is invertible, there exists an inverse matrix A^{-1} which is also invertible. Multiply equation (6.13) through by A^{-1} to obtain

$$\begin{aligned} A^{-1}\big(c_1 A\vec{x}_1 + c_2 A\vec{x}_2 + \cdots + c_k A\vec{x}_k\big) &= A^{-1}\vec{0}_n \\ c_1 A^{-1}A\vec{x}_1 + c_2 A^{-1}A\vec{x}_2 + \cdots + c_k A^{-1}A\vec{x}_k &= \vec{0}_n \\ c_1 I_n\vec{x}_1 + c_2 I_n\vec{x}_2 + \cdots + c_k I_n\vec{x}_k &= \vec{0}_n \\ c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_k\vec{x}_k &= \vec{0}_n. \end{aligned} \tag{6.14}$$

But equation (6.14) is a linear dependence relation for $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ contrary to it being a linearly independent set. Hence $\{A\vec{x}_1, A\vec{x}_2, \ldots, A\vec{x}_k\}$ must be linearly independent. □

Proof. (Of Theorem 6.3.2) Suppose A and B are similar $n \times n$ matrices, and let C be an invertible matrix such that $B = C^{-1}AC$. Then note that $B - \lambda I_n = C^{-1}AC - \lambda I_n$. We can write $I_n = C^{-1}I_nC$ and factor C^{-1} on the left side and C on the right side to obtain

$$\begin{aligned} B - \lambda I_n &= C^{-1}AC - \lambda C^{-1}I_nC \\ &= C^{-1}(AC - \lambda I_nC) \\ &= C^{-1}(A - \lambda I_n)C \end{aligned} \tag{6.15}$$

Equation 6.15 shows that $B - \lambda I_n$ and $A - \lambda I_n$ are similar matrices. Applying Theorem 6.3.1, they have the same determinant. That is,

$$P_B(\lambda) = \det(B - \lambda I_n) = \det(A - \lambda I_n) = P_A(\lambda).$$

Since the eigenvalues and their algebraic multiplicities are completely determine by the characteristic polynomial, A and B have the same eigenvalues with the same algebraic multiplicities. To demonstrate that the geometric multiplicity of each eigenvalue is the same, suppose λ_0 is an eigenvalue of A and let $\{\vec{x}_1, \ldots, \vec{x}_k\}$ be a basis for $E_A(\lambda_0)$, the eigenspace of A corresponding to the eigenvalue λ_0. From the similarity relationship $B = C^{-1}AC$, we have $BC^{-1} = C^{-1}A$. For each vector in our basis for the corresponding eigenspace, we have

$$BC^{-1}\vec{x}_i = C^{-1}A\vec{x}_i = C^{-1}\left(\lambda_0\vec{x}_i\right) = \lambda_0 C^{-1}\vec{x}_i. \tag{6.16}$$

Equation (6.16) shows that $C^{-1}\vec{x}_i$ is an eigenvector of B corresponding to the eigenvalue λ_0. As a basis, the set $\{\vec{x}_1, \ldots, \vec{x}_k\}$ is linearly independent, and as C^{-1} is invertible, Lemma 6.3.1 guarantees that $\{C^{-1}\vec{x}_1, \ldots, C^{-1}\vec{x}_k\}$ is linearly independent. Note that since A is also similar to B, we can construct an analogous argument to show that if $\vec{y}_i$ is any eigenvector for B corresponding to the eigenvalue λ_0, then $C\vec{y}_i$ is an eigenvector of A corresponding to λ_0. Hence $\{C^{-1}\vec{x}_1, \ldots, C^{-1}\vec{x}_k\}$ is a basis for the eigenspace $E_B(\lambda_0)$ of B corresponding to λ_0 consisting of the same number of basis elements as the basis for $E_A(\lambda_0)$. We conclude that the geometric multiplicity of λ_0 as an eigenvalue of B is the same as its geometric multiplicity as an eigenvalue of A. □

Remark 6.3.1. *Theorems 6.3.1 and 6.3.2 tell us that having the same determinant, characteristic equation and eigenvalues is a necessary consequence of being similar. However, none of these shared features is sufficient to conclude that two matrices are similar. That is, a pair of matrices may have the same characteristic polynomial but not be similar matrices. (The pair of matrices that feature in Exercises 6.2.4 and 6.2.5 are an example of matrices with the same characteristic polynomial that are not similar matrices. This is evidenced by the fact that the geometric multiplicity of the eignenvalue* 3 *is not the same for both matrices.)*

We are particularly interested in whether a given matrix A is similar to a diagonal matrix D. If this is the case, then Theorem 6.3.2 indicates that this diagonal matrix would have to have the eigenvalues of A as its diagonal entries (since the eigenvalues of any diagonal matrix are its diagonal entries, and A and D would have to have the same eigenvalues). Suppose our matrix A has n not necessarily distinct real eigenvalues $\lambda_1, \ldots \lambda_n$, whose geometric multiplies sum to n. Then our matrix A gives rise to an eigenbasis, $\mathcal{E}_A = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\}$, for R^n. With this set of n linearly independent vectors, we can construct an invertible matrix C having the eigenbasis elements as its column vectors, $\mathrm{Col}_i(C) = \vec{x}_i$. Now we consider the product AC. Focusing on the i^{th} column of this product, note that

$$\mathrm{Col}_i(AC) = A\,\mathrm{Col}_i(C) = A\vec{x}_i = \lambda_i\vec{x}_i = \lambda_i\,\mathrm{Col}_i(C). \tag{6.17}$$

Let $D = [d_{ij}]$ be the diagonal matrix defined by

$$d_{ii} = \lambda_i, \quad \text{and} \quad d_{ij} = 0 \quad \text{for} \quad i \neq j.$$

That is, $\text{Col}_i(D) = \lambda_i \vec{e}_i$, where as usual, $\vec{e}_i$ is the standard unit vector in R^n having a 1 in the i^{th} entry and zero everywhere else. Then the product CD will satisfy

$$\text{Col}_i(CD) = C\,\text{Col}_i(D) = C\lambda_i\vec{e}_i = \lambda_i C\vec{e}_i = \lambda_i\,\text{Col}_i(C). \tag{6.18}$$

(Here we've used the useful fact in equation (3.13) from Section 3.6.) Equations (6.17) and (6.18) show that each column of AC is equal to the corresponding column of CD. It follows that

$$CD = AC, \quad \text{i.e.,} \quad D = C^{-1}AC.$$

This wonderful observation provides a condition on diagonalizability as well as a formulation for the necessary invertible matrix.

Theorem 6.3.3. *Let A be an $n \times n$ matrix. Then A is diagonalizable if and only if A has n linearly independent eigenvectors. Moreover, if A is diagonalizable, then there exists a diagonal matrix D such that $D = C^{-1}AC$ where the columns of the invertible matrix C are the vectors in an eigenbasis, $\mathcal{E}_A$, for the matrix A, and the diagonal entries of the matrix D are the eigenvalues of A.*

Proof. Half of the proof of Theorem 6.3.3 is given in the construction preceding the theorem statement. There, we showed that if A has n linearly independent eigenvectors, then A is diagonalizable. Now, suppose that A is diagonalizable so that there exists a diagonal matrix $D = [d_{ij}]$ and invertible matrix C such that $D = C^{-1}AC$. Then as before, we have $CD = AC$. Note that making use of the fact that $\text{Col}_i(D) = d_{ii}\vec{e}_i$, the i^{th} column of CD is

$$\text{Col}_i(CD) = C\,\text{Col}_i(D) = Cd_{ii}\vec{e}_i = d_{ii}C\vec{e}_i = d_{ii}\,\text{Col}_i(C). \tag{6.19}$$

The i^{th} column of the product AC

$$\text{Col}_i(AC) = A\,\text{Col}_i(C). \tag{6.20}$$

Comparing equations (6.19) and (6.20), we see that

$$A\,\text{Col}_i(C) = d_{ii}\,\text{Col}_i(C).$$

Since C is invertible, each column vector of C is nonzero. Hence each diagonal entry d_{ii} is an eigenvalue of A with corresponding eigenvector $\text{Col}_i(C)$. The column vectors of the invertible matrix C are linearly independent, so we conclude that A has n linearly independent eigenvectors. □

Example 6.3.1. *Show that the matrix A is diagonalizable by finding a diagonal matrix D and invertible matrix C such that $D = C^{-1}AC$.*

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

First, we find the characteristic polynomial $\det(A - \lambda I_3)$*. Taking a cofactor expansion across the third row (to take advantage of the zeros there)*

$$\begin{aligned} \det\left(\begin{bmatrix} 4-\lambda & 7 & 1 \\ 1 & -2-\lambda & 1 \\ 0 & 0 & 1-\lambda \end{bmatrix}\right) &= (1-\lambda)\big((4-\lambda)(-2-\lambda) - 1(7)\big) \\ &= (1-\lambda)(\lambda^2 - 2\lambda - 15) \\ &= (1-\lambda)(\lambda-5)(\lambda+3) \end{aligned}$$

So A has three eigenvalues $\lambda_1 = 1$, $\lambda_2 = 5$ and $\lambda_3 = -3$. Next, we find a basis for each eigenspace. Since we have three distinct eigenvalues, we're guaranteed to find an eigenbasis with three linearly independent eigenvectors. For $\lambda_1 = 1$,

$$\left[A - 1I_3|\vec{0}_3\right] \quad \xrightarrow{rref} \quad \left[\begin{array}{ccc|c} 1 & 0 & \frac{5}{8} & 0 \\ 0 & 1 & -\frac{1}{8} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Eigenvectors will have the form $\vec{x}_1 = t\left\langle -\frac{5}{8}, \frac{1}{8}, 1\right\rangle$, $t \in R$. For $\lambda_2 = 5$,

$$\left[A - 5I_3|\vec{0}_3\right] \quad \xrightarrow{rref} \quad \left[\begin{array}{ccc|c} 1 & -7 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Eigenvectors will have the form $\vec{x}_2 = t\langle 7, 1, 0\rangle$, $t \in R$. And for $\lambda_3 = -3$,

$$\left[A - (-3)I_3|\vec{0}_3\right] \quad \xrightarrow{rref} \quad \left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Eigenvectors will have the form $\vec{x}_3 = t\langle -1, 1, 0\rangle$, $t \in R$. *We need one of each for the matrix* C, *let's call these vectors* $\vec{v}_1$, $\vec{v}_2$ *and* $\vec{v}_3$. *For the first eigenvalue, let's choose* $t = 8$ *to get* $\vec{v}_1 = \langle -5, 1, 8\rangle$. *We can choose* $t = 1$ *for the other two to get* $\vec{v}_2 = \langle 7, 1, 0\rangle$ *and* $\vec{v}_3 = \langle -1, 1, 0\rangle$. *Letting these be the columns of* C *in this order,*

$$C = \begin{bmatrix} -5 & 7 & -1 \\ 1 & 1 & 1 \\ 8 & 0 & 0 \end{bmatrix} \quad \textit{with inverse} \quad C^{-1} = \frac{1}{16}\begin{bmatrix} 0 & 0 & 2 \\ 2 & 2 & 1 \\ -2 & 14 & -3 \end{bmatrix}.$$

Then the diagonal matrix

$$D = C^{-1}AC = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -3 \end{bmatrix}.$$

We might note that the matrix C constructed in Example 6.3.1 is not unique. In addition to selecting which eigenvectors would be used as the columns of C, we selected the order in which the columns would appear. We could have made other choices. Given the construction in the proof of Theorem 6.3.3, it should be clear that the order in which the eigenvalues appear on the diagonal of D corresponds to the order in which the eigenvectors appear as columns in C. So for the preceding example, selecting the eigenvectors for 1, 5 and -3 in this order resulted in the eigenvalues appearing in D in this order. We could agree to a specific order, such as numerically increasing, but there is no universal convention for this sort of matrix decomposition.

Exercise 6.3.2. *For each matrix, either diagonalize the matrix (i.e., identify the diagonal matrix* D *and invertible matrix* C*) or show that the matrix is not diagonalizable.*

1. $A = \begin{bmatrix} -4 & 7 \\ -2 & 5 \end{bmatrix}$

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

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

4. $B = \begin{bmatrix} -2 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 6 \end{bmatrix}$

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

Exercise 6.3.3.

1. *Find a* 3×3 *matrix* A *having eigenvalues* $L = \{1, -4, 5\}$ *and for which* $\mathcal{E}_A = \{\langle 1, 1, 3\rangle, \langle 1, 1, -3\rangle, \langle 0, -1, -2\rangle\}$ *is an eigenbasis.*

2. *Is your answer* A *in part 1. above unique? That is, can you find another* 3×3 *matrix having eigenvalues* $L = \{1, -4, 5\}$ *and eigenbasis* $\mathcal{E}_A = \{\langle 1, 1, 3\rangle, \langle 1, 1, -3\rangle, \langle 0, -1, -2\rangle\}$*?*

Because the algebraic multiplicity of an eigenvalues is greater than or equal to its geometric multiplicity, we have the following result.

Theorem 6.3.4. *If* A *is an* $n \times n$ *matrix with* n *distinct eigenvalues, then* A *is diagonalizable.*

Theorem 6.3.4 provides a sufficient condition for diagonalizability. As you've seen in Exercise 6.3.1 it is not necessary that a matrix has n distinct eigenvalues. A matrix with fewer than n eigenvalues may or may not be diagonalizable and must be considered on a case by case basis.

An advantage of diagonal matrices is the ease with which successive matrix multiplication, including computing successive powers, can be done. Suppose we have an $n \times n$ matrix A that we wish to evaluate powers of, say A^2, A^3, A^4, and so forth. Even a 2×2 inspires the use of technology. Consider the relatively simple matrix

$$A = \begin{bmatrix} -4 & 3 \\ -6 & 5 \end{bmatrix}.$$

Note that

$$A^2 = \begin{bmatrix} -4 & 3 \\ -6 & 5 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ -6 & 5 \end{bmatrix} = \begin{bmatrix} -2 & 3 \\ -6 & 7 \end{bmatrix}$$

$$A^3 = A^2 A = \begin{bmatrix} -2 & 3 \\ -6 & 7 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ -6 & 5 \end{bmatrix} = \begin{bmatrix} -10 & 9 \\ -18 & 17 \end{bmatrix}$$

$$A^4 = A^3A = \begin{bmatrix} -10 & 9 \\ -18 & 17 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ -6 & 5 \end{bmatrix} = \begin{bmatrix} -14 & 15 \\ -30 & 31 \end{bmatrix}$$

Such computations readily become tiresome, especially if we desire much larger powers, A^5, A^{10}, A^{25}. Compare to the ease with which we can compute power of the matrix

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

In fact,

$$D^{10} = \begin{bmatrix} 2^{10} & 0 \\ 0 & (-1)^{10} \end{bmatrix} = \begin{bmatrix} 1024 & 0 \\ 0 & 1 \end{bmatrix}.$$

It isn't even necessary to pass through each of the powers 2 through 9. If our matrix A is diagonalizable, we can take advantage of the similar diagonal matrix. First, note that if A and B are similar matrices, then A^2 and B^2 are also similar. To confirm this, suppose $B = C^{-1}AC$, for some invertible matrix C. Then note that

$$B^2 = (C^{-1}AC)^2 = (C^{-1}AC)(C^{-1}AC) = C^{-1}A(CC^{-1})AC =$$

$$= C^{-1}AI_nAC = C^{-1}AAC = C^{-1}A^2C.$$

Not only are A^2 and B^2 similar, they share the same transformation as A and B. More generally, we have the following theorem that can be proven by induction.

Theorem 6.3.5. *If A and B are similar matrices and C is an invertible matrix such that $B = C^{-1}AC$, then for each integer $n \geq 1$, A^n and B^n are similar and $B^n = C^{-1}A^nC$.*

Example 6.3.2. *Evaluate A^{10} where $A = \begin{bmatrix} -4 & 3 \\ -6 & 5 \end{bmatrix}$.*

If A is diagonalizable, we can compute this with two matrix multiplications and one matrix inversion as opposed to nine matrix multiplications. The eigenvalues of A are found to be $\lambda_1 = 2$ and $\lambda_2 = -1$. With two distinct eigenvalues, we are assured that A is diagonalizable. Associated eigenvectors are $\vec{x}_1 = \langle 1, 2\rangle$ and $\vec{x}_2 = \langle 1, 1\rangle$, so $D = C^{-1}AC$ where

$$D = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}, \quad \text{and} \quad C = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}.$$

The inverse matrix

$$C^{-1} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}.$$

Now, $A^n = CD^nC^{-1}$*, so*

$$\begin{aligned} A^{10} &= \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2^{10} & 0 \\ 0 & (-1)^{10} \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} -1(2^{10}) & 1(2^{10}) \\ 2(-1)^{10} & -1(-1)^{10} \end{bmatrix} \\ &= \begin{bmatrix} -2^{10} + 2(-1)^{10} & 2^{10} - (-1)^{10} \\ -2(2^{10}) + 2(-1)^{10} & 2(2^{10}) - (-1)^{10} \end{bmatrix} \\ &= \begin{bmatrix} -1022 & 1023 \\ -2046 & 2047 \end{bmatrix} \end{aligned}$$

6.4 Linear Transformations and Change of Basis

Theorem 5.2.1 provides a construction for the $n \times n$ standard matrix A associated with a linear transformation $T : R^n \to R^n$. We recall that the column vectors for the matrix are the images of the standard basis vectors under T, $\mathrm{Col}_j(A) = T(\vec{e}_j)$. There is a subtle, yet critical, bias built into that construction, namely that the vectors $\vec{x}$ in the domain as well as their images $T(\vec{x})$ are to be represented by their coordinates relative to the standard basis $\mathcal{E} = \{\vec{e}_1, \ldots, \vec{e}_n\}$. What if we desire some other basis, for example one in which the corresponding matrix is diagonal?

In Section 4.3.1, we defined coordinate vectors for subspaces of R^n (which can include all of R^n). Here, let's consider an alternative basis for R^n, say $\mathcal{C} = \{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}$. Given a vector $\vec{x}$ in R^n, we can consider its coordinate vector relative to this new basis

$$[\vec{x}]_{\mathcal{C}} = \langle \alpha_1, \alpha_2, \ldots, \alpha_n \rangle,$$

where the entries are the unique coefficients of $\vec{x}$ in terms of the basis $\mathcal{C}$,

$$\vec{x} = \alpha_1 \vec{c}_1 + \alpha_2 \vec{c}_2 + \cdots + \alpha_n \vec{c}_n. \tag{6.21}$$

As a linear combination of vectors, we can rephrase equation (6.21) as a matrix-vector product

$$\vec{x} = C[\vec{x}]_{\mathcal{C}},$$

where C is the $n\times n$ matrix whose columns are the basis vectors, $\text{Col}_i(C) = \vec{c}_i$. Given that the columns of C form a basis for R^n, the matrix C is invertible. This provides us with a way to translate back and forth between the standard basis and the new basis,

$$\vec{x} = C[\vec{x}]_{\mathcal{C}}, \quad \text{and} \quad [\vec{x}]_{\mathcal{C}} = C^{-1}\vec{x}. \tag{6.22}$$

We can call the matrix C a **change of basis matrix** for the basis $\mathcal{C}$.

Example 6.4.1. *Consider the ordered basis $\mathcal{C} = \{\langle 1,2\rangle, \langle 1,1\rangle\}$ of R^2.*

1. *Identify the change of basis matrix C and its inverse C^{-1}.*
2. *Find the coordinate vectors relative to the basis $\mathcal{C}$ for the following vectors.*
   (a) $\vec{x} = \langle 1,0\rangle$
   (b) $\vec{y} = \langle 3,-2\rangle$
   (c) $\vec{z} = \langle 5,1\rangle$
3. *Find the representation relative to the standard basis for the vectors having the given coordinate vectors relative to the basis $\mathcal{C}$.*
   (a) $[\vec{u}]_{\mathcal{C}} = \langle 1,0\rangle$
   (b) $[\vec{v}]_{\mathcal{C}} = \langle 3,-2\rangle$
   (c) $[\vec{w}]_{\mathcal{C}} = \langle 5,1\rangle$

Solutions

1. *The change of basis matrix* $C = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$. *Its inverse is* $C^{-1} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}$.
2. *To find the coordinate vectors relative to the basis $\mathcal{C}$, we use the relationship $[\vec{x}]_{\mathcal{C}} = C^{-1}\vec{x}$.*
   (a) $[\vec{x}]_{\mathcal{C}} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} \langle 1,0\rangle = \langle -1,2\rangle$
   (b) $[\vec{y}]_{\mathcal{C}} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} \langle 3,-2\rangle = \langle -5,8\rangle$

(c) $[\vec{z}]_{\mathcal{C}} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} \langle 5, 1 \rangle = \langle -4, 9 \rangle$

3. *To find the vectors in terms of the standard basis having the given coordinate vectors relative to* $\mathcal{C}$*, we use the relationship* $\vec{x} = C[\vec{x}]_{\mathcal{C}}$*.*

   (a) $\vec{u} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \langle 1, 0 \rangle = \langle 1, 2 \rangle$

   (b) $\vec{v} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \langle 3, -2 \rangle = \langle 1, 4 \rangle$

   (c) $\vec{w} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \langle 5, 1 \rangle = \langle 6, 11 \rangle$

Exercise 6.4.1. *Consider the ordered basis* $\mathcal{C} = \{\langle 1, 1 \rangle, \langle -1, 5 \rangle\}$ *of* R^2*.*

1. *Identify the change of basis matrix* C *and its inverse* C^{-1}*.*

2. *Find the coordinate vectors relative to the basis* $\mathcal{C}$ *for the following vectors.*

   (a) $\vec{x} = \langle 1, 1 \rangle$

   (b) $\vec{y} = \langle -1, 5 \rangle$

   (c) $\vec{z} = \langle 0, 1 \rangle$

3. *Find the representation relative to the standard basis for the vectors having the given coordinate vectors relative to the basis* $\mathcal{C}$*.*

   (a) $[\vec{u}]_{\mathcal{C}} = \langle 1, 1 \rangle$

   (b) $[\vec{v}]_{\mathcal{C}} = \langle -1, 5 \rangle$

   (c) $[\vec{w}]_{\mathcal{C}} = \langle 0, 1 \rangle$

If $T : R^n \to R^n$ is a linear transformation and $\mathcal{C} = \{\vec{c}_1, \ldots, \vec{c}_n\}$ is a basis of R^n, we can express a vector $\vec{x}$ as well as $T(\vec{x})$, its image under T, in terms of their coordinate vectors relative to the basis $\mathcal{C}$. The transformation that maps a vector $\vec{x}$ in R^n to a coordinate vector $[\vec{x}]_{\mathcal{C}}$ in R^n is a linear transformation[4].

[4]This is rather obvious in R^n since mapping a vector to a coordinate vector is matrix-vector multiplication, but Lemma 4.8.1 in Section 4.8 ensures that the mapping to coordinate vectors is linear in general vector spaces as well.

So the composition of a linear transformation T with a coordinate mapping will be a linear transformation. This suggests that there is some matrix B associated with the transformation in the alternative coordinate system,

$$\big[T(\vec{x})\big]_{\mathcal{C}} = B[\vec{x}]_{\mathcal{C}}.$$

We could call this new matrix the $\mathcal{C}$-matrix for the linear transformation T. How is the $\mathcal{C}$-matrix, B, related to the standard matrix A, where $T(\vec{x}) = A\vec{x}$?

The relationship between the $\mathcal{C}$-matrix, B, and the standard matrix, A, can be deduced by applying the linear transformation T as well as the change of basis transformation from equation (6.22). For some vector $\vec{x}$ in R^n, we can express its image $T(\vec{x})$ under T in terms of its coordinate vector relative to the basis $\mathcal{C}$. If C is the change of basis matrix, then for each vector $\vec{y}$ in R^n, we have $[\vec{y}]_{\mathcal{C}} = C^{-1}\vec{y}$. So the coordinate vector for $T(\vec{x})$ relative to $\mathcal{C}$ is

$$\big[T(\vec{x})\big]_{\mathcal{C}} = C^{-1}T(\vec{x}). \tag{6.23}$$

Now, since $T(\vec{x}) = A\vec{x}$ with A the standard matrix for T, we have

$$\big[T(\vec{x})\big]_{\mathcal{C}} = C^{-1}\left(A\vec{x}\right) = C^{-1}A\vec{x}. \tag{6.24}$$

Finally, from equation (6.22) we can replace $\vec{x}$ with $C[\vec{x}]_{\mathcal{C}}$ to arrive at

$$\big[T(\vec{x})\big]_{\mathcal{C}} = C^{-1}A\left(C[\vec{x}]_{\mathcal{C}}\right) = \left(C^{-1}AC\right)[\vec{x}]_{\mathcal{C}}. \tag{6.25}$$

We see that the $\mathcal{C}$-matrix that we called B above is similar to the standard matrix A. Specifically,

$$B = C^{-1}AC$$

where C is the change of basis matrix for our alternative basis $\mathcal{C}$. If A is diagonalizable, we can use an eigenbasis to formulate the linear transformation in terms of a diagonal matrix. But we should note that the above derivation is not restricted to diagonal matrices. That is, we didn't insist that the basis $\mathcal{C}$ has to be an eigenbasis or that the matrix B must be diagonal. We can use this to express a linear transformation from R^n into R^n in terms of any basis for R^n.

Example 6.4.2. *Let $T : R^2 \to R^2$ be the linear transformation defined by $T(\vec{x}) = A\vec{x}$ where $A = \begin{bmatrix} -4 & 3 \\ -6 & 5 \end{bmatrix}$. This is the diagonalizable matrix from Example 6.3.2 where we found that $D = C^{-1}AC$ with*

$$D = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}, \quad \textit{and} \quad C^{-1} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}.$$

This tells us that for the basis $\mathcal{C} = \{\langle 1,2\rangle, \langle 1,1\rangle\}$ of R^2, the $\mathcal{C}$-matrix for T is the diagonal matrix D. Let's confirm that

$$\big[T(\vec{x})\big]_{\mathcal{C}} = B[\vec{x}]_{\mathcal{C}}$$

for the vectors $\vec{e}_1$ and $\vec{e}_2$.

First, let's find the coordinate vectors for $\vec{e}_1$ and $\vec{e}_2$ relative to the basis $\mathcal{C}$. These are

$$[\vec{e}_1]_{\mathcal{C}} = C^{-1}\vec{e}_1 = \langle -1, 2\rangle, \quad \text{and} \quad [\vec{e}_2]_{\mathcal{C}} = C^{-1}\vec{e}_2 = \langle 1, -1\rangle.$$

The images of $\vec{e}_1$ and $\vec{e}_2$ under T are

$$T(\vec{e}_1) = A\vec{e}_1 = \langle -4, -6\rangle, \quad \text{and} \quad T(\vec{e}_2) = A\vec{e}_2 = \langle 3, 5\rangle.$$

Equation (6.25) indicates that we should be able to compute $\big[T(\vec{e}_i)\big]_{\mathcal{C}}$ using the diagonal $\mathcal{C}$-matrix. This gives

$$\big[T(\vec{e}_1)\big]_{\mathcal{C}} = D[\vec{e}_1]_{\mathcal{C}} = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix} \langle -1, 2\rangle = \langle -2, -2\rangle, \quad \text{and}$$

$$\big[T(\vec{e}_2)\big]_{\mathcal{C}} = D[\vec{e}_2]_{\mathcal{C}} = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix} \langle 1, -1\rangle = \langle 2, 1\rangle.$$

Alternatively, we can find the coordinate vectors $\big[T(\vec{e}_i)\big]_{\mathcal{C}}$ by applying the inverse of the change of basis matrix to the images $T(\vec{e}_i)$ we already found. Using this approach gives

$$\big[T(\vec{e}_1)\big]_{\mathcal{C}} = C^{-1}T(\vec{e}_1) = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} \langle -4, -6\rangle = \langle -2, -2\rangle, \quad \text{and}$$

$$\big[T(\vec{e}_2)\big]_{\mathcal{C}} = C^{-1}T(\vec{e}_2) = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} \langle 3, 5\rangle = \langle 2, 1\rangle.$$

The two approaches to computing the coordinate vectors, $\big[T(\vec{e}_i)\big]_{\mathcal{C}}$, for the images yields the same results, as they should. (Granted, this was a rather tedious, perhaps a less than practical, exercise. The example is just intended to illustrate equation (6.25).)

Exercise 6.4.2. *Let $T : R^2 \to R^2$ be the linear transformation defined by $T(\vec{x}) = A\vec{x}$ where $A = \begin{bmatrix} -2 & 6 \\ -2 & 5 \end{bmatrix}$. Find a basis $\mathcal{C}$ of R^2 such that the $\mathcal{C}$-matrix of T is diagonal. Find the $\mathcal{C}$-matrix.*

Exercise 6.4.3. *A matrix A is called* **symmetric** *if $A = A^T$. It is known that symmetric matrices are always diagonalizable. Moreover, the eigenvectors for distinct eigenvalues are orthogonal. That is, a symmetric matrix has an eigenbasis of mutually orthogonal vectors. Let $T : R^3 \to R^3$ be the linear transformation defined by $T(\vec{x}) = A\vec{x}$ for the matrix A given below. Find a basis $\mathcal{C}$ of R^3 such that the $\mathcal{C}$-matrix of T is diagonal, and confirm that the basis elements are orthogonal. Find the $\mathcal{C}$-matrix.*

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

Exercise 6.4.4. *Let $T : R^2 \to R^2$ be the shear transformation such that $T(\vec{e}_1) = \vec{e}_1 - 2\vec{e}_2$ and $T(\vec{e}_2) = \vec{e}_2$ (so T leaves $\vec{e}_2$ fixed). Determine whether there is a basis $\mathcal{C}$ of R^2 such that the $\mathcal{C}$-matrix of T is diagonal. If so, find the diagonal matrix.*

6.5 Additional Exercises

(Jump to Solutions)

1. If $A = [a_{11}]$ is a 1×1 matrix, we define its determinant to be $\det(A) = a_{11}$. Use this definition to show that the determinant of a 2×2 matrix from Definition 6.1.1 is the same as a cofactor expansion

$$\det(A) = \sum_{j=1}^{2} (-1)^{1+j} a_{1j} \det(A_{1j}).$$

   (The point of this exercises is to show that the determinant of a 2×2 really is computed using the same cofactor expansion used for larger matrices.)

2. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Suppose A has two (not necessarily distinct) eigenvalues λ_1 and λ_2. Show that

$$a + d = \lambda_1 + \lambda_2 \quad \text{and} \quad \det(A) = \lambda_1 \lambda_2.$$

   (Hint: The characteristic polynomial must factor as $P_A(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)$. Compare this to P_A obtained in the usual way.)

3. Give a coherent argument that if $A = [a_{ij}]$ is an $n \times n$ triangular matrix, the eigenvalues of A are its diagonal entries, a_{ii}.

4. Suppose A is an invertible matrix. Show that $\det(A^{-1}) = (\det(A))^{-1}$. That is, show that the determinant of A^{-1} is the reciprocal of the determinant of A.

5. For the matrix A, evaluate $\det(A)$. Find all of the eigenvalues of A and show that $\det(A)$ is equal to the product of the eigenvalues of A.

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

6. Suppose the $n \times n$ matrix A has n not necessarily distinct real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Show that $\det(A) = \lambda_1\lambda_2 \cdots \lambda_n$, that is, the determinant of A is the product of its eigenvalues.
(Hint: The characteristic polynomial can be written as a product of linear factors $P_A(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda)$. How are $\det(A)$ and $P_A(0)$ related?)

7. Suppose A is an $n \times n$ invertible matrix and λ_0 is a non-zero eigenvalue of A. Show that $\dfrac{1}{\lambda_0}$ is an eigenvalue of A^{-1}.

8. Suppose A is a 5×5 matrix with characteristic polynomial

$$P_A(\lambda) = (2 - \lambda)^2(4 - \lambda)(-1 - \lambda)(6 - \lambda).$$

For each question, either provide a short answer or explain why it is not possible to answer.

(a) Is A invertible?
(b) Evaluate $\det(A - 2I_5)$
(c) Is A diagonalizable?
(d) Is there a nonzero vector $\vec{x}$ in R^5 such that $A\vec{x} = -\vec{x}$?
(e) What is $\det(A)$?
(f) Is $\det(A - 5I_5) = 0$?
(g) Is there an eigenbasis of R^5 for A?

9. Suppose A is a 5×5 matrix with characteristic polynomial

$$P_A(\lambda) = -\lambda(1-\lambda)(-1-\lambda)(2-\lambda)(7-\lambda).$$

For each question, either provide a short answer or explain why it is not possible to answer.

   (a) Is A invertible?
   (b) Is $A - I_5$ invertible?
   (c) Is A diagonalizable?
   (d) Is there a nonzero vector $\vec{x}$ in R^5 such that $A\vec{x} = \vec{x}$?
   (e) What is $\det(A)$?
   (f) Is $A - 5I_5$ invertible?
   (g) Is there an eigenbasis of R^5 for A?

10. Find a 3×3 matrix A having eigenvalues $L = \{2, -1, 3\}$ and eigenbasis $\mathcal{E}_A = \{\langle 1, 0, 1\rangle, \langle -2, 1, 0\rangle, \langle 3, 1, 2\rangle\}$.

11. Prove Theorem 6.3.5. That is, show that if A and B are similar, then for positive integer n, A^n and B^n are also similar. (Hint: using induction.)

12. Suppose A and B are similar, invertible matrices. Show that A^{-1} and B^{-1} are similar and that A^T and B^T are similar.

13. (Involves calculus) An interesting use of diagonalization arises in the solution of linear systems of differential equations. We know, for example, that the simple differential equation $\frac{dy}{dt} = ay$, with a a constant, has family of solutions $y(t) = e^{at}y_0$ where y_0 is a scalar (it is the value of $y(t)$ when $t = 0$). We can formulate a vector version of this simple equation with $\vec{y}(t) = \langle x(t), y(t)\rangle$, a vector valued function of t. The derivative is taken entry-wise, $\frac{d\vec{y}}{dt} = \left\langle \frac{dx}{dt}, \frac{dy}{dt}\right\rangle$. If A is a 2×2 matrix, we can consider the vector differential equation

$$\frac{d\vec{y}}{dt} = A\vec{y},$$

and propose a solution analogous to the scalar version, $\vec{y}(t) = e^{At}\vec{y}_0$. This requires giving meaning to an exponential e^{At} when A is a matrix.

We can turn to a series representation. Recall that the exponential e^x can be expressed in terms of the series

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

This suggests a way to give meaning to a matrix exponential. We can define

$$e^{At} = I_n + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \cdots = \sum_{n=0}^{\infty} \frac{t^n}{n!}A^n.$$

If D is a diagonal matrix, $D = \begin{bmatrix} d_{11} & 0 \\ 0 & d_{22} \end{bmatrix}$, then we can get a nice form for the exponential of the matrix,

$$e^{Dt} = \begin{bmatrix} e^{d_{11}t} & 0 \\ 0 & e^{d_{22}t} \end{bmatrix}.$$

Glossing over some technical issues, we can show that if $D = C^{-1}AC$, then $e^{At} = Ce^{Dt}C^{-1}$. Determine the matrix exponential e^{At} if $A = \begin{bmatrix} -2 & 6 \\ -2 & 5 \end{bmatrix}$. (Note this is the matrix from Exercise 6.4.2.) Show that $e^{A(0)} = I_2$, that is, when $t = 0$, the matrix exponential is the identity (this is analogous to the fact that $e^0 = 1$).

Chapter 7

Orthogonality & Projections

Consider a W pound truck parked on ramp that makes an angle θ with respect to the horizontal as shown in Figure 7.1, and suppose we are interested in the force required by the brakes to maintain its position without rolling down the ramp. We can represent the weight of the truck as a vector $\vec{w} = -W\vec{e}_2$ in R^2 based on a simple, local model of the Earth's gravitational field. The force exerted by the brakes will be parallel to the incline, so it is desirable to express the weight using a coordinate system with axes parallel and perpendicular to the ramp. The braking force is then the component of $\vec{w}$ that is parallel to the ramp.

The point of this scenario is not to solve this simple physics problem but rather to motivate the orthogonal projection of a vector onto some subspace; in this case, a line in R^2. In general, suppose H is a subspace of R^n and $\vec{y}$ is a vector that is not in H. We can ask whether there is a vector $\vec{y}_H$ that is an element of H that is *closest to* $\vec{y}$ as illustrated in Figure 7.2. Like the truck's weight and the ramp, we can consider such a vector $\vec{y}_H$ as a projection onto the subspace H such that the difference $\vec{y} - \vec{y}_H$ is perpendicular to H.

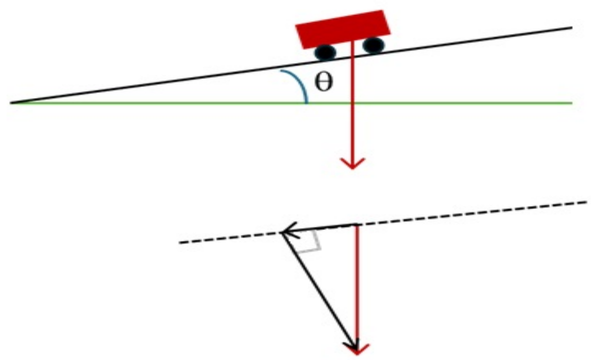


Figure 7.1: The weight of an object, a vertical vector, on an incline can be decomposed as the sum of a vector parallel to the incline and a vector orthogonal to the incline.

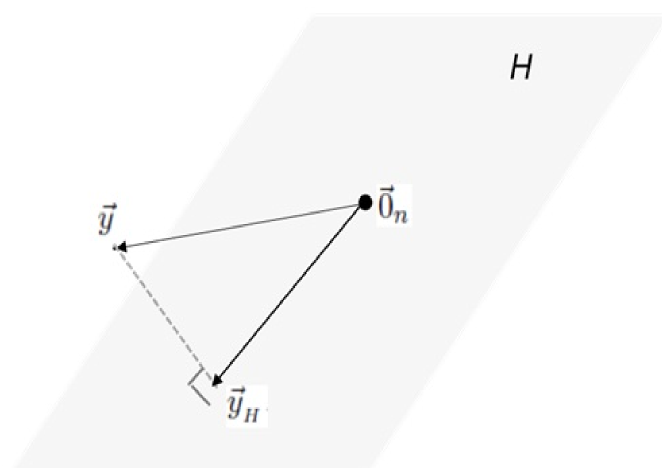


Figure 7.2:

7.1 Orthogonal Sets & Bases

In Chapter 1, we defined *orthogonality* of a pair of vectors in terms of the dot product, and we saw that for nonzero vectors in R^n, orthogonality has a geometric interpretation; nonzero orthogonal vectors are perpendicular. Now, we wish to consider sets of vectors that are mutually orthogonal.

Definition 7.1.1. *Let $k \geq 2$ and let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ be a set of nonzero vectors in R^n. We will call S and* **orthogonal set** *if for each $i, j = 1, \ldots, k$*

$$\vec{v}_i \cdot \vec{v}_j = 0, \quad \textit{whenever} \quad i \neq j.$$

A simple and familiar example of an orthogonal set is the standard basis $\{\vec{e}_1, \ldots, \vec{e}_n\}$ in R^n. This corresponds to our geometric sense of a set of mutually perpendicular coordinate axes. The standard unit vectors are often desirable because of the ease with which we can express vectors as a linear combination of them. We will find that orthogonal sets generally provide for a computationally simple approach to linear combinations. This is true even when the vectors in an orthogonal set are more exotic than standard unit vectors.

Example 7.1.1. *Consider the vectors $\vec{v}_1 = \langle 3, 0, -3, 1\rangle$, $\vec{v}_2 = \langle 2, 1, 1, -3\rangle$ and $\vec{v}_3 = \langle 1, 5, 2, 3\rangle$ in R^4. Determine whether the set $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is an orthogonal set.*

Solution: *To determine whether the set is orthogonal, we must consider three dot products $\vec{v}_1 \cdot \vec{v}_2$, $\vec{v}_1 \cdot \vec{v}_3$, and $\vec{v}_2 \cdot \vec{v}_3$. S is orthogonal if each of these is zero.*

$$\begin{aligned}
\vec{v}_1 \cdot \vec{v}_2 &= 3(2) + 0(1) + (-3)(1) + 1(-3) = 0\\
\vec{v}_1 \cdot \vec{v}_3 &= 3(1) + 0(5) + (-3)(2) + 1(3) = 0\\
\vec{v}_2 \cdot \vec{v}_3 &= 2(1) + 1(5) + 2(1) + (-3)(3) = 0
\end{aligned}$$

Each pair of distinct vectors in S is orthogonal, and we conclude that S is an orthogonal set.

Example 7.1.2. *Consider the vectors $\vec{v}_1 = \langle 1, -1, 3\rangle$, $\vec{v}_2 = \langle -1, 2, 1\rangle$ and $\vec{v}_3 = \langle 6, 3, -1\rangle$ in R^3. Determine whether the set $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is an orthogonal set.*

Solution: *As in the last example, we must consider three dot products* $\vec{v}_1 \cdot \vec{v}_2$, $\vec{v}_1 \cdot \vec{v}_3$, *and* $\vec{v}_2 \cdot \vec{v}_3$. S *is orthogonal if each of these is zero.*

$$\begin{aligned}\vec{v}_1 \cdot \vec{v}_2 &= 1(-1) + (-1)(2) + 3(1) = 0\\ \vec{v}_1 \cdot \vec{v}_3 &= 1(6) + (-1)(3) + 3(-1) = 0\\ \vec{v}_2 \cdot \vec{v}_3 &= -1(6) + 2(3) + 1(-1) = -1\end{aligned}$$

While $\vec{v}_1$ *is orthogonal to both* $\vec{v}_2$ *and* $\vec{v}_3$, $\vec{v}_2$ *and* $\vec{v}_3$ *are not orthogonal. So our conclusion is that* S *is not an orthogonal set.*

Exercise 7.1.1. *Let* $S = \{\langle 1, -1, 3\rangle, \langle -1, 2, 1\rangle, \langle 7, 4, -1\rangle\}$. *Show that* S *is an orthogonal set.*

One immediate consequence of a set of vectors being orthogonal is that such a set is necessarily linearly independent. While the zero vector $\vec{0}_n$ is orthogonal to every vector in R^n, Definition 7.1.1 specifies that $\vec{0}_n$ is not an element of an orthogonal set.

Theorem 7.1.1. *Let* $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$, *where* $k \geq 2$, *be an orthogonal set of vectors in* R^n. *Then* S *is linearly independent.*

Proof. Consider the homogeneous vector equation

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k = \vec{0}_n. \tag{7.1}$$

To demonstrate that S is linearly independent, we must show that $c_i = 0$ for each $i = 1, \ldots, k$. To show that $c_1 = 0$, we can take the dot product of each side of equation (7.1) with $\vec{v}_1$. Making use of the algebraic properties of the dot product, we have

$$\begin{aligned}\vec{v}_1 \cdot (c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k) &= \vec{v}_1 \cdot \vec{0}_n\\ c_1\vec{v}_1 \cdot \vec{v}_1 + c_2\vec{v}_1 \cdot \vec{v}_2 + \cdots + c_k\vec{v}_1 \cdot \vec{v}_k &= 0\end{aligned}$$

Now, each of $\vec{v}_1 \cdot \vec{v}_i = 0$ for $i = 2, \ldots, k$, and $\vec{v}_1 \cdot \vec{v}_1 = \|\vec{v}_1\|^2$. So this reduces to

$$c_1\|\vec{v}_1\|^2 = 0.$$

Since $\vec{v}_1$ is a nonzero vector, $\|\vec{v}_1\|^2$ is some positive number, and we see that $c_1 = 0$, necessarily. We can take this same approach to isolate each of the weights, c_i. If we take the dot product of each side of equation (7.1) with $\vec{v}_i$, we obtain an equation $c_i\|\vec{v}_i\|^2 = 0$ from which we conclude that $c_i = 0$. Equation 7.1 has only the trivial solution, $c_1 = c_2 = \cdots = c_k = 0$, and S is linearly independent. □

An immediate consequence of Theorem 7.1.1 is the following.

Corollary 7.1.1. *If S is an orthogonal set of vectors in R^n, then S is a basis for the subspace* Span(S) *of R^n.*

As one might expect, we will call such a basis, an orthogonal basis.

Definition 7.1.2. *Let H be a subspace of R^n. An* **orthogonal basis** *for H is a basis that is an orthogonal set.*

As noted, the standard unit vectors, $\{\vec{e}_1, \ldots, \vec{e}_n\}$, form an orthogonal basis of R^n. This example of an orthogonal basis has the additional property that each vector is a unit vector. The term **orthonormal** captures these two properties.

Definition 7.1.3. *An* **orthonormal set** *is a set $S = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ of unit vectors that is an orthogonal set. An* **orthonormal basis** *of a subspace of R^n is a basis that is an orthonormal set.*

In the proof of Theorem 7.1.1, we saw that the mutual orthogonality of the vectors can be exploited to isolate a single term in a linear combination. In that proof, we were focused on a homogeneous equation, but this same approach can be used to determine the weights for any linear combination. Take for example the set $S = \{\langle 1,-1,3\rangle, \langle -1,2,1\rangle, \langle 7,4,-1\rangle\}$. In Exercise 7.1.1, you confirmed that this is an orthogonal set. Given that S contains three linearly independent vectors, $\dim(\text{Span}(S)) = 3$, and as R^3 is the only 3-dimensional subspace of R^3, we can say that S is an orthogonal basis for R^3. Hence every vector in R^3 can be expressed as a linear combination of the vectors in the set S. Suppose we wish to express the vector $\vec{x} = \langle 1,2,3\rangle$ in terms of the basis S,

$$\vec{x} = \langle 1,2,3\rangle = c_1\langle 1,-1,3\rangle + c_2\langle -1,2,1\rangle + c_3\langle 7,4,-1\rangle. \tag{7.2}$$

There are various approaches to this task. For example, we can treat it as a system of linear equations and perform row reduction on the augmented matrix $\left[\begin{array}{rrr|r} 1 & -1 & 7 & 1 \\ -1 & 2 & 4 & 2 \\ 3 & 1 & -1 & 3 \end{array}\right]$. Alternatively, we can introduce a change of basis matrix to find the coordinate vector for $\vec{x}$ relative to the basis S:

$$\langle c_1, c_2, c_3 \rangle = [\vec{x}]_S = \begin{bmatrix} 1 & -1 & 7 \\ -1 & 2 & 4 \\ 3 & 1 & -1 \end{bmatrix}^{-1} \vec{x}.$$

The orthogonality of S provides a far less computationally intensive approach to obtaining these weights. Note that if we take the dot product of each side of equation 7.2 with $\langle 1, -1, 3 \rangle$ and use the orthogonality, we quickly isolate the coefficient c_1 of $\langle 1, -1, 3 \rangle$.

$$\langle 1, -1, 3 \rangle \cdot \langle 1, 2, 3 \rangle = c_1 \|\langle 1, -1, 3 \rangle\|^2 + c_2(0) + c_3(0).$$

Hence

$$c_1 = \frac{\langle 1, -1, 3 \rangle \cdot \langle 1, 2, 3 \rangle}{\|\langle 1, -1, 3 \rangle\|^2} = \frac{8}{11}.$$

Similar calculations give

$$c_2 = \frac{\langle -1, 2, 1 \rangle \cdot \langle 1, 2, 3 \rangle}{\|\langle -1, 2, 1 \rangle\|^2} = \frac{6}{6} = 1,$$

and

$$c_3 = \frac{\langle 7, 4, -1 \rangle \cdot \langle 1, 2, 3 \rangle}{\|\langle 7, 4, -1 \rangle\|^2} = \frac{12}{66} = \frac{2}{11}.$$

We find that

$$\langle 1, 2, 3 \rangle = \frac{8}{11} \langle 1, -1, 3 \rangle + \langle -1, 2, 1 \rangle + \frac{2}{11} \langle 7, 4, -1 \rangle.$$

Theorem 7.1.2 generalizes this observation.

Theorem 7.1.2. *Let H be a subspace of R^n and $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ be an orthogonal basis of H. Then each vector $\vec{x}$ in H can be expressed as*

$$\vec{x} = c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_k \vec{v}_k, \quad \textit{where} \quad c_i = \frac{\vec{v}_i \cdot \vec{x}}{\|\vec{v}_i\|^2}.$$

Exercise 7.1.2. *Let* $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ *where*

$$\vec{v}_1 = \langle 3, 0, -3, 1\rangle, \quad \vec{v}_2 = \langle 2, 1, 1, -3\rangle, \quad \textit{and} \quad \vec{v}_3 = \langle 1, 5, 2, 3\rangle.$$

In Example 7.1.1, we determined that S *is an orthogonal basis for* $\operatorname{Span}(S)$. *Use the formula for the weights from Theorem 7.1.2 to express* $\vec{x} = \langle 3, 3, -2, 4\rangle$ *as a linear combination of the elements of* S *and confirm that your solution is correct.*

We can restate the result of Theorem 7.1.2 in the case of an orthonormal basis. Specifically, if $S = \{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_k\}$ is an orthonormal basis for a subspace H of R^n and $\vec{x}$ is any element of H, we have

$$\vec{x} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k, \quad \text{where} \quad c_i = \vec{u}_i \cdot \vec{x}.$$

The coefficient formulation is simplified because $\|\vec{u}_i\|^2 = 1$. We recall from Chapter 1 that given a nonzero vector $\vec{x}$ in R^n, the direction vector, $\vec{x}_U = \frac{1}{\|\vec{x}\|}\vec{x}$, is a unit vector in the direction of $\vec{x}$. Given an orthogonal set $\{\vec{v}_1, \dots, \vec{v}_k\}$, we can readily construct an orthonormal set $\{\vec{u}_1, \dots, \vec{u}_k\}$ by setting

$$\vec{u}_i = \frac{1}{\|\vec{v}_i\|}\vec{v}_i.$$

This process of scaling vectors to obtain unit vectors is often called **normalizing**, hence the term "orthonormal."

Exercise 7.1.3. *Show that the set* $\{\langle 2, 2, 1\rangle, \langle -2, 1, 2\rangle, \langle 1, -2, 2\rangle\}$ *is an orthogonal basis for* R^3 *and find an associated orthonormal basis by normalizing the vectors.*

7.2 Orthogonal Projections

Returning to the question suggested at the beginning of this chapter, given a subspace H of R^n and a vector $\vec{y}$ that is not necessarily in H, we can seek a vector $\vec{y}_H$ in H such that $\vec{y} = \vec{y}_H + \vec{z}$ where the vector $\vec{z}$ is orthogonal to every vector in H. The vector $\vec{y}_H$ can be thought of as the *part* of the vector $\vec{y}$ that is in the subspace H, and we refer to this as a **projection**.

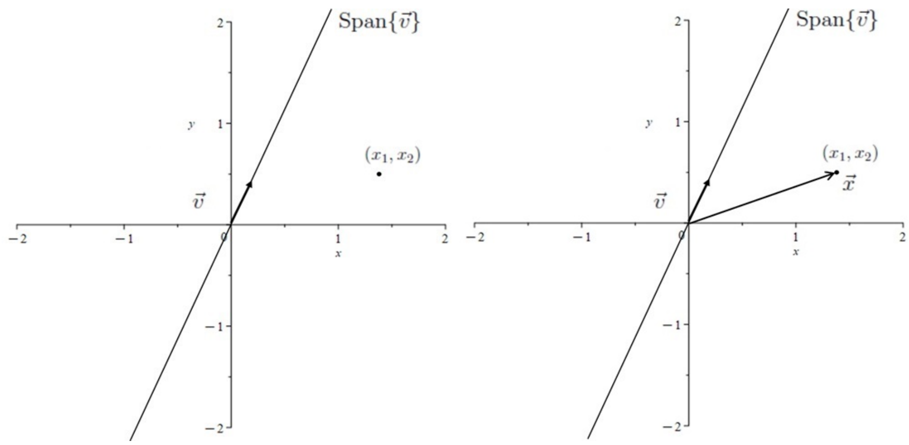


Figure 7.3: Left: A subspace Span$\{\vec{v}\}$ of R^2, and a point (x_1, x_2) not on H. Right: The standard representation of $\vec{x} = \langle x_1, x_2 \rangle$.

7.2.1 Projection Onto a Vector

We begin by considering the projection of one vector onto another vector or, equivalently, onto a one-dimensional subspace of R^n—i.e., a line through the origin. R^2 is a convenient setting since we can visualize things graphically, however, our construction extends readily to R^n (even if we don't have nice pictures). If $\vec{v}$ is a nonzero vector in R^2, then the subspace $H = \text{Span}\{\vec{v}\}$ can be associated with a line that is parallel to $\vec{v}$ and passes through the origin. Consider a point (x_1, x_2) in R^2 that is not necessarily on the line H as shown on the left in Figure 7.3. What point on the line H is closest to the point (x_1, x_2), and what is the distance between this point and the line?

If $\vec{x} = \langle x_1, x_2 \rangle$, then the standard representation of $\vec{x}$ will terminate at the point (x_1, x_2) as shown on the right in Figure 7.3. Now, we want to express $\vec{x}$ as the sum

$$\vec{x} = \vec{x}_{_H} + \vec{z},$$

where $\vec{x}_{_H}$ is in H and $\vec{z}$ is orthogonal to H. As an element of H, we know that $\vec{x}_{_H} = k\vec{v}$ for some scalar k. Hence

$$\vec{x} = k\vec{v} + \vec{z}, \tag{7.3}$$

and since $\vec{z}$ should be perpendicular[1] to H, $\vec{z} \cdot \vec{v} = 0$. To determine the scalar

[1]More precisely, $\vec{z}$ is orthogonal to $\vec{v}$ which includes the case that $\vec{z} = \vec{0}_2$ as would be the case when $\vec{x}$ is already in H.

k, we take the dot product of each side of equation (7.3) with $\vec{v}$. Making use of the orthogonality, we have

$$\vec{x} \cdot \vec{v} = (k\vec{v} + \vec{z}) \cdot \vec{v} = k\vec{v} \cdot \vec{v} + \vec{z} \cdot \vec{v} = k\|\vec{v}\|^2 + 0.$$

We see that the scalar

$$k = \frac{\vec{x} \cdot \vec{v}}{\|\vec{v}\|^2}, \tag{7.4}$$

and the vector

$$\vec{x}_H = \frac{\vec{x} \cdot \vec{v}}{\|\vec{v}\|^2}\vec{v}. \tag{7.5}$$

We recognize the expression in equation (7.4) from the formulas for the weights appearing in Theorem 7.1.2. The remaining vector, what we labeled $\vec{z}$ in equation (7.3), is $\vec{z} = \vec{x} - \vec{x}_H$. We can confirm that this vector is in fact orthogonal to every vector in $H = \text{Span}\{\vec{v}\}$. Each vector in H has the form $c\vec{v}$ for some scalar c. Using equation (7.5) and the algebraic properties of the dot product,

$$\begin{aligned}
(c\vec{v}) \cdot \vec{z} = c\,(\vec{v} \cdot \vec{z}) = c\,(\vec{v} \cdot (\vec{x} - \vec{x}_H)) &= c\left(\vec{v} \cdot \vec{x} - \vec{v} \cdot \left(\frac{\vec{x} \cdot \vec{v}}{\|\vec{v}\|^2}\vec{v}\right)\right) \\
&= c\left(\vec{v} \cdot \vec{x} - \left(\frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|^2}\right)\vec{v} \cdot \vec{v}\right) \\
&= c\left(\vec{v} \cdot \vec{x} - \left(\frac{\vec{x} \cdot \vec{v}}{\cancel{\|\vec{v}\|^2}}\right)\cancel{\|\vec{v}\|^2}\right) \\
&= c\,(\vec{v} \cdot \vec{x} - \vec{x} \cdot \vec{v}) \\
&= c\,(0) \\
&= 0.
\end{aligned}$$

Hence $\vec{x} - \vec{x}_H$ is orthogonal to H. The vector $\vec{x}_H$ derived in equation (7.5) is called a vector projection.

Definition 7.2.1. *Let $\vec{v}$ be a nonzero vector in R^n and let $\vec{x}$ be a vector in R^n. The* **vector projection of $\vec{x}$ onto $\vec{v}$** *is denoted $\text{proj}_{\vec{v}}\,\vec{x}$ and defined by*

$$\text{proj}_{\vec{v}}\,\vec{x} = \frac{\vec{x} \cdot \vec{v}}{\|\vec{v}\|^2}\vec{v}. \tag{7.6}$$

Remark 7.2.1. *There are two key vectors in a vector projection; the vector being projected (here $\vec{x}$) and the nonzero vector being projected onto (here $\vec{v}$). The vector* $\operatorname{proj}_{\vec{v}} \vec{x}$ *is a scalar multiple of the vector $\vec{v}$, hence it is an element of* $\operatorname{Span}\{\vec{v}\}$. *In the notation presented in Definition 7.6, the nonzero vector being projected onto is written as a subscript of "*proj*" where as the vector being projected is written as an argument of "*proj*".*

Remark 7.2.2. *The vector projection given in Definition 7.6 can also be called the* **vector projection of $\vec{x}$ onto the subspace** $\operatorname{Span}\{\vec{v}\}$. *If $H = \operatorname{Span}\{\vec{v}\}$, then the notation* $\operatorname{proj}_{\vec{v}} \vec{x}$, *can be replaced with* $\operatorname{proj}_{\operatorname{Span}\{\vec{v}\}} \vec{x}$ *or* $\operatorname{proj}_H \vec{x}$.

Back to our example in R^2, if $\operatorname{proj}_{\vec{v}} \vec{x} = \langle a, b\rangle$, then (a, b) is the point on the line $\operatorname{Span}\{\vec{v}\}$ that is closest to the point (x_1, x_2). The distance between the point (x_1, x_2) and this line is the magnitude, $\|\vec{z}\|$, of the orthogonal difference, $\vec{x} - \vec{x}_H$, as shown on the right side of Figure 7.4.

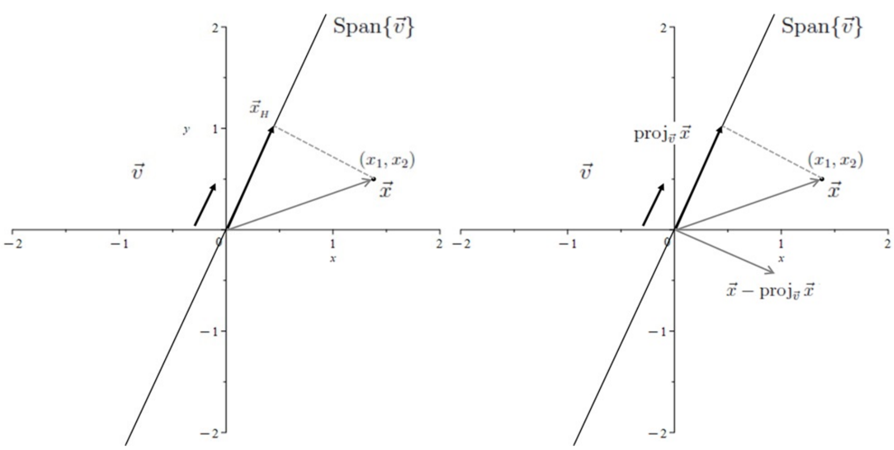


Figure 7.4: Left: The terminal point of the standard representation of $\vec{x}_H$ is the point on $\operatorname{Span}\{\vec{v}\}$ closest to (x_1, x_2). Right: The distance between the point (x_1, x_2) and the line $\operatorname{Span}\{\vec{v}\}$ is $\|\vec{x} - \operatorname{proj}_{\vec{v}} \vec{x}\|$, the magnitude of vector $\vec{x} - \operatorname{proj}_{\vec{v}} \vec{x}$.

Example 7.2.1. *Let L be the line $3x + 2y = 0$, and consider the point $P = (-5, 2)$. What is the point on L that is closest to P? What is the distance between L and P?*

Solution: *Note that the point $B = (-2, 3)$ is on L, so the vector $\vec{v} = \overrightarrow{OB} = \langle -2, 3\rangle$ is parallel to L. Let $\vec{x} = \overrightarrow{OP} = \langle -5, 2\rangle$. To find the point on L closest to P, we compute the projection $\text{proj}_{\vec{v}}\,\vec{x}$. Using the formula from Definition 7.2.1*

$$\text{proj}_{\vec{v}}\,\vec{x} = \frac{\vec{x}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v} = \frac{16}{13}\langle -2, 3\rangle.$$

So the point on L closest to P is $\left(-\frac{32}{13}, \frac{48}{13}\right)$. To determine the distance between P and L, we need the difference,

$$\vec{x} - \text{proj}_{\vec{v}}\,\vec{x} = \langle -5, 2\rangle - \frac{16}{13}\langle -2, 3\rangle = \left\langle \frac{97}{13}, -\frac{22}{13}\right\rangle.$$

The distance is the magnitude of this vector,

$$\textit{Distance from } P \textit{ to } L = \left\|\left\langle \frac{97}{13}, -\frac{22}{13}\right\rangle\right\| = \sqrt{\frac{761}{13}} \approx 7.65.$$

You might wonder where the choice of the vector $\vec{v} = \langle -2, 3\rangle$ came from in Example 7.2.1. In general, we can select any two points, say $A = (a_1, a_2)$ and $B = (b_1, b_2)$, on the line L to identify a vector $\vec{v} = \overrightarrow{AB}$ parallel to L. Does the resulting projection change if we choose a different vector to represent the direction of the line L? (If it did, that would certainly call our solution into question!) Note that for any nonzero vector $\vec{v}$ and any vector $\vec{x}$

$$\frac{\vec{x}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v} = \left(\frac{\vec{x}\cdot\vec{v}}{\|\vec{v}\|}\right)\frac{\vec{v}}{\|\vec{v}\|} = \left(\vec{x}\cdot\frac{\vec{v}}{\|\vec{v}\|}\right)\frac{\vec{v}}{\|\vec{v}\|} = (\vec{x}\cdot\vec{v}_U)\,\vec{v}_U, \tag{7.7}$$

where as was defined in Section 1.3, $\vec{v}_U = \dfrac{\vec{v}}{\|\vec{v}\|}$ is the direction vector of the vector $\vec{v}$. So the projection formula presented in Definition 7.2.1 depends on the direction of the vector $\vec{v}$ being projected onto, but it is independent of the magnitude of $\vec{v}$. We have been considering examples in R^2, but the construction in equation (7.7) places no restriction on the number of entries in the vectors $\vec{x}$ and $\vec{v}$ which can be elements of R^n for any $n \geq 2$.

The scalar $\dfrac{\vec{x}\cdot\vec{v}}{\|\vec{v}\|}$ is typically called the **scalar component** of the vector $\vec{x}$ in the direction of the vector $\vec{v}$. Its value is indicative of the relationship between the vectors $\vec{x}$ and $\vec{v}$. In particular, if the vectors $\vec{x}$ and $\vec{v}$ are perpendicular, then they are orthogonal and

$$\text{proj}_{\vec{v}}\,\vec{x} = \frac{\vec{x}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v} = \frac{0}{\|\vec{v}\|^2}\vec{v} = \vec{0}_n.$$

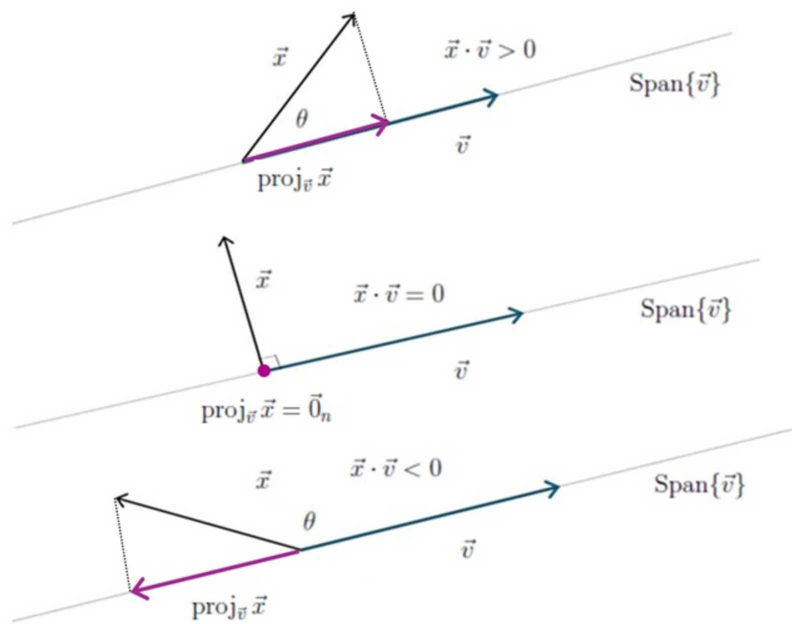


Figure 7.5: Graphical representation of the projection of a vector $\vec{x}$ onto a vector $\vec{v}$, equivalently the subspace Span$\{\vec{v}\}$. When $\vec{x}\cdot\vec{v}>0$ (top), $\text{proj}_{\vec{v}}\,\vec{x}$ is in the direction of $\vec{v}$. When $\vec{x}\cdot\vec{v}=0$ (middle), $\text{proj}_{\vec{v}}\,\vec{x}=\vec{0}_n$. And when $\vec{x}\cdot\vec{v}<0$ (bottom), $\text{proj}_{\vec{v}}\,\vec{x}$ is parallel to, but in the opposite direction from $\vec{v}$.

If $\vec{x}$ is a nonzero vector that is parallel to $\vec{v}$, then the projection of $\vec{x}$ onto $\vec{v}$ is simply $\vec{x}$ (see Exercise 7.2.3). If $\vec{x}$ is nonzero and is neither parallel nor perpendicular to $\vec{v}$, then the sign of $\dfrac{\vec{x}\cdot\vec{v}}{\|\vec{v}\|}$ tells us about the nature of the angle between[2] the vectors $\vec{x}$ and $\vec{v}$. Since $\|\vec{v}\|>0$, the sign of the scalar component is determined by the sign of $\vec{x}\cdot\vec{v}$. If $\vec{x}\cdot\vec{v}>0$, then the angle formed by $\vec{x}$ and $\vec{v}$ is acute, and if $\vec{x}\cdot\vec{v}<0$, the angle is obtuse. These cases are illustrated in Figure 7.5.

Exercise 7.2.1. *Consider the parallel vectors* $\vec{v}_1=\langle -2,3\rangle$ *and* $\vec{v}_2=\langle 4,-6\rangle$*,*

[2]In R^n for $n\geq 3$, two nonzero and nonparallel vectors determine a plane, and the angle being referenced is the angle between the vectors in this plane.

and let $\vec{x} = \langle -5, 2\rangle$. *Show that*

$$\operatorname{proj}_{\vec{v}_1} \vec{x} = \operatorname{proj}_{\vec{v}_2} \vec{x}.$$

Exercise 7.2.2. *Find the point on the line* L *defined by* $4x - y = 0$ *closest to the point* $P = (6, 1)$. *What is the distance between the point* P *and the line* L*?*

Exercise 7.2.3.

1. *Let* $\vec{v} = \langle 1, -1, 2, -3\rangle$ *and* $\vec{x} = \langle 3, -3, 6, -9\rangle$. *Verify that* $\vec{x}$ *is parallel to* $\vec{v}$ *and that* $\operatorname{proj}_{\vec{v}} \vec{x} = \vec{x}$.

2. *Let* $\vec{v}$ *be any nonzero vector in* R^n. *Show that if* $\vec{x}$ *is any vector in* R^n *that is parallel to* $\vec{v}$, *then* $\operatorname{proj}_{\vec{v}} \vec{x} = \vec{x}$.

7.2.2 Projection Onto a Subspace

7.3 Additional Exercises

(Jump to Solutions)

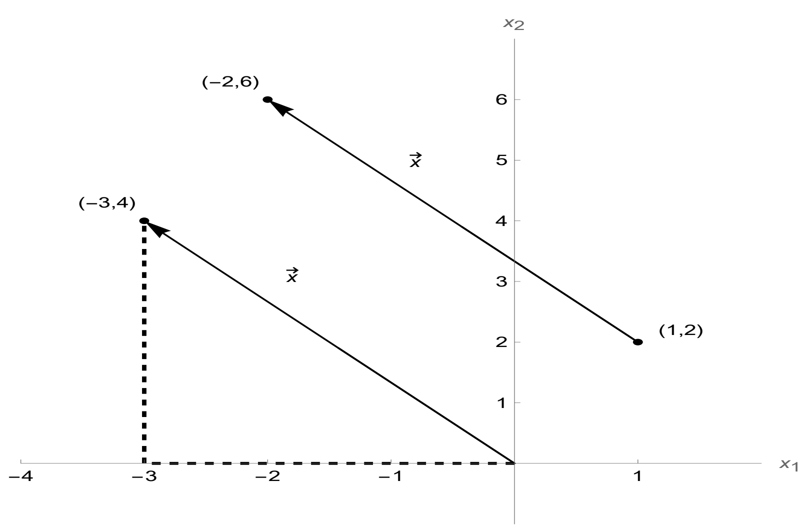
Appendix A

Answers and Solutions to Selected Exercises

A.1 Chapter 1 Exercises:

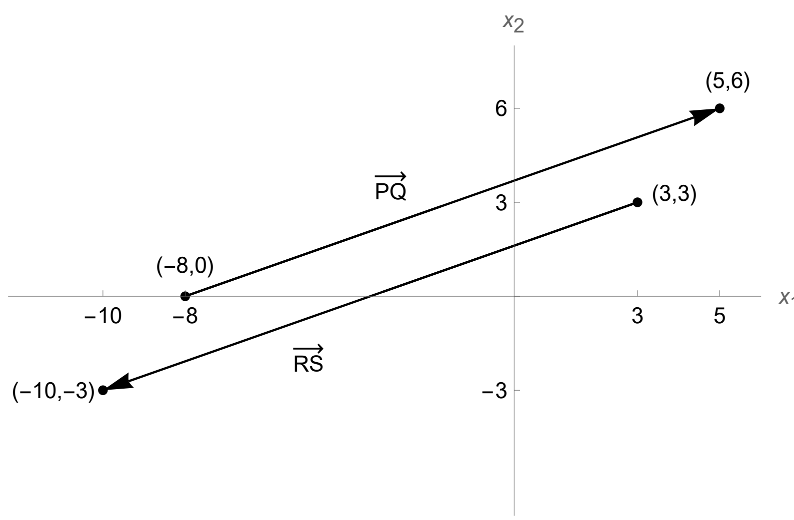
Exercise 1.1.1: Draw a picture of the standard representative of the vector $\vec{x} = \langle -3, 4 \rangle$. Then draw a picture of the representative of $\vec{x}$ that is based at the point $P = (1, 2)$. (To do this you will need to find the point Q such that $\overrightarrow{PQ}$ is a representative of $\vec{x}$.)

Solution:



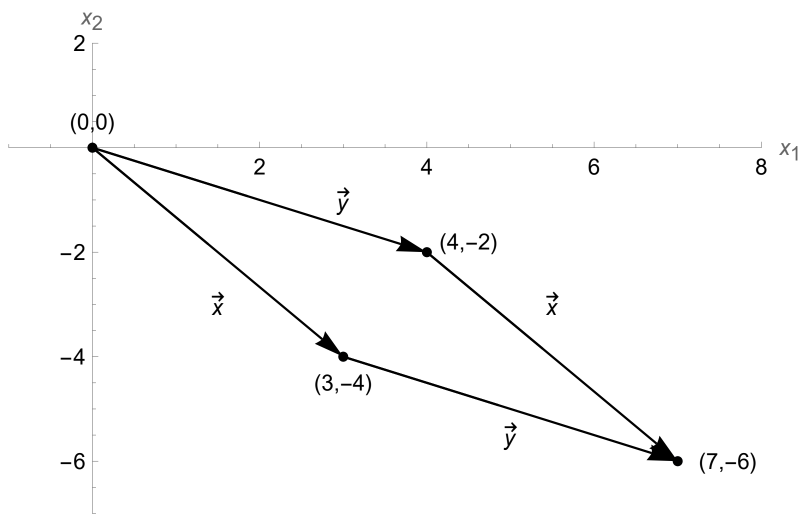
Exercise 1.1.3: In parts 1–5 below, four points (P, Q, R, and S) are given. Draw the directed line segments $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ and determine whether or not $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ represent the same vector. If they do not represent the same vector, then state whether this is because they don't have the same length or don't point in the same direction (or both).

Solution to number 4: The points given are $P = (-8, 0)$, $Q = (5, 6)\,, R = (3, 3)\,, S = (-10, -3)$ and we see that $\overrightarrow{PQ} = \langle 13, 6 \rangle$ and $\overrightarrow{RS} = \langle -13, -6 \rangle$. These vectors are not equal. They do have the same length but they do not point in the same direction. (They point in opposite directions.)



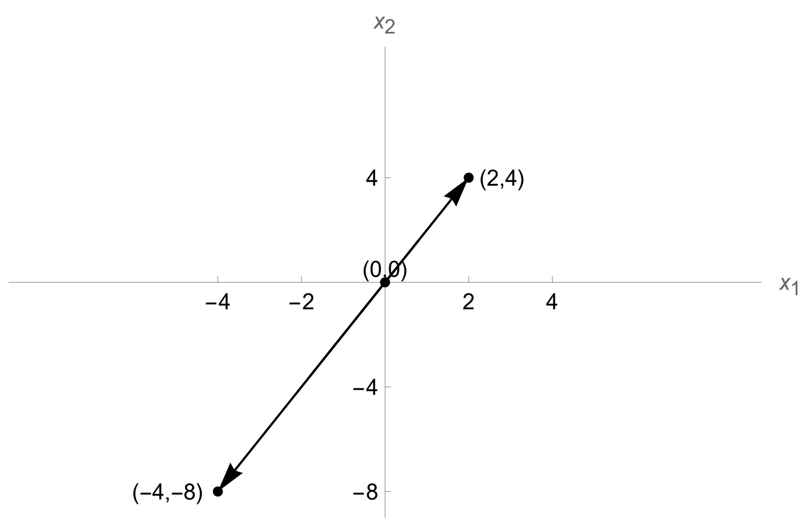
Exercise 1.1.4: For each pair of vectors, $\vec{x}$ and $\vec{y}$, given in parts 1–7, compute $\vec{x} + \vec{y}$ and then draw a picture to illustrate the Parallelogram Method of Vector Addition for $\vec{x} + \vec{y}$.

Solution to number 1: We are given $\vec{x} = \langle 3, -4 \rangle$ and $\vec{y} = \langle 4, -2 \rangle$. Thus $\vec{x} + \vec{y} = \langle 7, -6 \rangle$.



Exercise 1.1.7:

Solution to number 2: Since $\vec{x} = \langle 2, 4\rangle$, then $-2\vec{x} = \langle -4, -8\rangle$ Both of these vectors are pictured below. Note that $-2\vec{x}$ has double the length of $\vec{x}$ and points in the opposite direction.



Exercise 1.1.9:

Solution to number 1: Since $\vec{x} = \langle -4, -4\rangle$, $\vec{y} = \langle -3, -2\rangle$, $c = 1$, $d = 4$, we have

$$c\vec{x} = (1)\vec{x} = \langle -4, -4\rangle$$

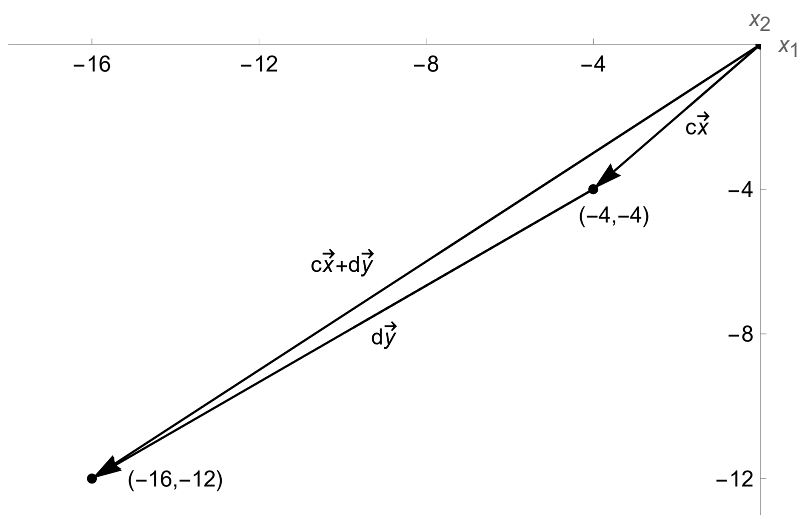
and

$$d\vec{y} = (4)\vec{y} = \langle -12, -8 \rangle .$$

Thus

$$c\vec{x} + d\vec{y} = \langle -16, -12 \rangle .$$

$c\vec{x}$, $d\vec{y}$, and $c\vec{x} + d\vec{y}$ are pictured below.



Exercise 1.1.11:
Solution to number 3:
The length of $\vec{x} = \langle -6, 4 \rangle$ is

$$\|\vec{x}\| = \sqrt{(-6)^2 + (4)^2} = \sqrt{52}.$$

Exercise 1.1.13:

Solution to number 3. Let us show that if $\vec{x}$ is any vector in R^2 and c is any scalar, then the length of $c\vec{x}$ is equal to the absolute value of c times the length of $\vec{x}$. In other words, let us show that

$$\|c\vec{x}\| = |c| \; \|\vec{x}\| .$$

To see this, let $\vec{x} = \langle x_1, x_2 \rangle$. Then $\|\vec{x}\| = \sqrt{x_1^2 + x_2^2}$ and $c\vec{x} = \langle cx_1, cx_2 \rangle$.

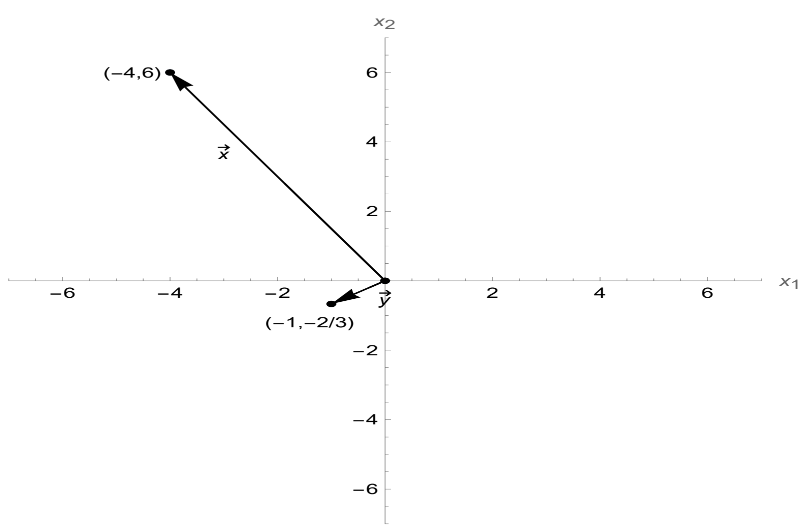
By computation, we see that

$$\begin{aligned}\|c\vec{x}\| &= \sqrt{(cx_1)^2 + (cx_2)^2} \\ &= \sqrt{c^2x_1^2 + c^2x_2^2} \\ &= \sqrt{c^2\left(x_1^2 + x_2^2\right)} \\ &= \sqrt{c^2}\sqrt{x_1^2 + x_2^2} \\ &= |c|\ \|\vec{x}\|\end{aligned}$$

Exercise 1.1.15:

Solutions for numbers 3 and 4:

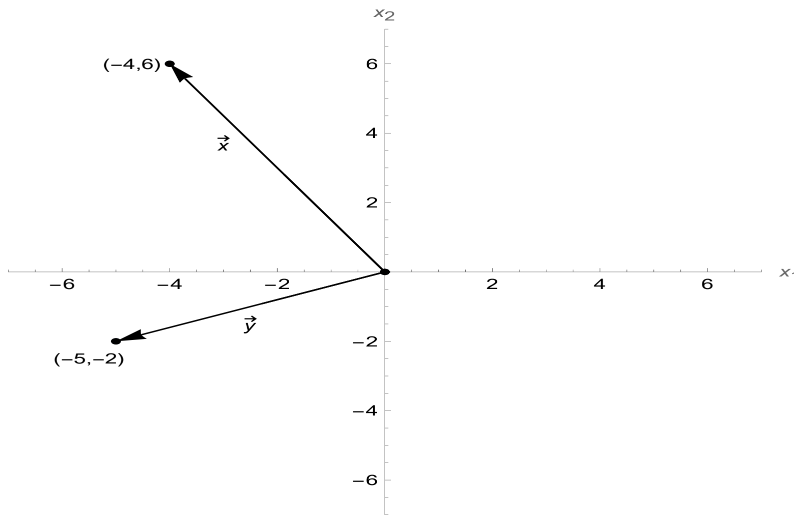
3) For $\vec{x} = \langle -4, 6\rangle$ and $\vec{y} = \left\langle -1, -\frac{2}{3}\right\rangle$, we have $\vec{x}\cdot\vec{y} = (-4)(-1) + (6)\left(-\frac{2}{3}\right) = 0$ and this tells us that $\vec{x}$ and $\vec{y}$ are orthogonal to each other. $\vec{x}$ and $\vec{y}$ are pictured below.



4) For $\vec{x} = \langle -4, 6\rangle$ and $\vec{y} = \langle -5, -2\rangle$, we have

$$\vec{x}\cdot\vec{y} = (-4)(-5) + (6)(-2) = 8 \neq 0.$$

This tells us that $\vec{x}$ and $\vec{y}$ are not orthogonal to each other. $\vec{x}$ and $\vec{y}$ are pictured below.



Exercise 1.1.17:
Solution for Number 2:
We will prove that if $\vec{x}$, $\vec{y}$ and $\vec{z}$ are any three vectors in R^2, then

$$\vec{x} \cdot (\vec{y} + \vec{z}) = \vec{x} \cdot \vec{y} + \vec{x} \cdot \vec{z}.$$

To do this, we write out $\vec{x}$, $\vec{y}$ and $\vec{z}$ in terms of their components

$$\begin{aligned} \vec{x} &= \langle x_1, x_2 \rangle \\ \vec{y} &= \langle y_1, y_2 \rangle \\ \vec{z} &= \langle z_1, z_2 \rangle, \end{aligned}$$

and note that

$$\begin{aligned} \vec{x} \cdot (\vec{y} + \vec{z}) &= \langle x_1, x_2 \rangle \cdot \langle y_1 + z_1, y_2 + z_2 \rangle \\ &= x_1 (y_1 + z_1) + x_2 (y_2 + z_2) \\ &= x_1 y_1 + x_1 z_1 + x_2 y_2 + x_2 z_2 \\ &= (x_1 y_1 + x_2 y_2) + (x_1 z_1 + x_2 z_2) \\ &= \vec{x} \cdot \vec{y} + \vec{x} \cdot \vec{z}. \end{aligned}$$

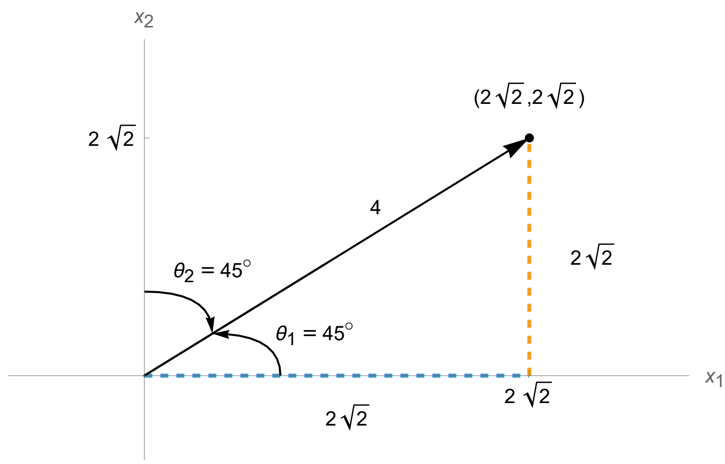
Exercise 1.1.19:
Solution for Number 1: Note that

$$\vec{x}_U = \langle \cos(45^\circ), \cos(45^\circ) \rangle = \left\langle \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right\rangle$$

and thus

$$\vec{x} = \|\vec{x}\|\,\vec{x}_U = 4\left\langle \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right\rangle = \left\langle 2\sqrt{2}, 2\sqrt{2} \right\rangle.$$

See picture below.



Solution for Number 3:

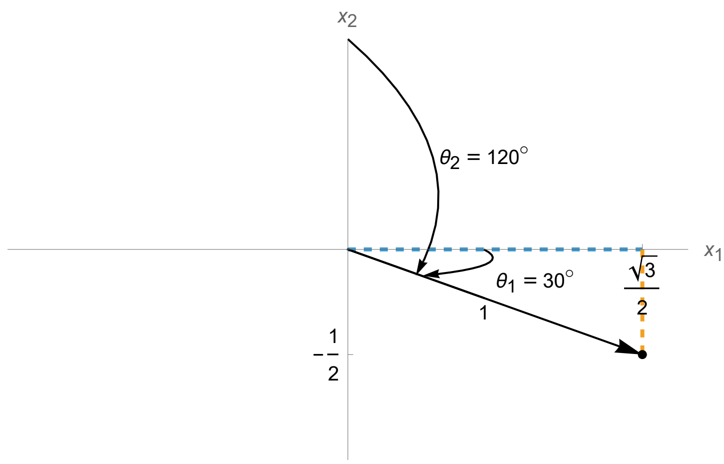
Note that

$$\vec{x}_U = \langle \cos(30°), \cos(120°) \rangle = \left\langle \frac{\sqrt{3}}{2}, -\frac{1}{2} \right\rangle$$

and thus

$$\vec{x} = \|\vec{x}\|\,\vec{x}_U = 1\left\langle \frac{\sqrt{3}}{2}, -\frac{1}{2} \right\rangle = \left\langle \frac{\sqrt{3}}{2}, -\frac{1}{2} \right\rangle.$$

See picture below.



Exercise 1.1.21:
The magnitude of $\vec{x}$ is

$$\|\vec{x}\| = \sqrt{3^2 + 6^2} = \sqrt{45} = 3\sqrt{5}$$

and the direction cosines of $\vec{x}$ are

$$\frac{x_1}{\|\vec{x}\|} = \frac{3}{3\sqrt{5}} = \frac{1}{\sqrt{5}}$$
$$\frac{x_2}{\|\vec{x}\|} = \frac{6}{3\sqrt{5}} = \frac{2}{\sqrt{5}}.$$

Thus the direction vector of $\vec{x}$ is

$$\vec{x}_U = \left\langle \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right\rangle.$$

Since $\vec{y}$ has magnitude 5 and points in the same direction as $\vec{x}$, then

$$\vec{y} = 5\vec{x}_U = 5\left\langle \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right\rangle = \left\langle \sqrt{5}, 2\sqrt{5} \right\rangle.$$

Exercise 1.1.22 Show that $\text{dist}(\vec{y}, \vec{x})$ is equal to $\text{dist}(\vec{x}, \vec{y})$ for any pair of vectors $\vec{x}$ and $\vec{y}$.

Answer: Letting $\vec{x} = \langle x_1, x_2 \rangle$ and $\vec{y} = \langle y_1, y_2 \rangle$,

$$\begin{aligned}
\text{dist}(\vec{x}, \vec{y}) &= \|\vec{x} - \vec{y}\| \\
&= \|\langle x_1 - y_1, x_2 - y_2 \rangle\| \\
&= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \\
&= \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2} \\
&= \|\langle y_1 - x_1, y_2 - x_2 \rangle\| \\
&= \|\vec{y} - \vec{x}\| \\
&= \text{dist}(\vec{y}, \vec{x}).
\end{aligned}$$

Exercise 1.1.23 Find the distance between each set of vectors.

1. $\vec{x} = \langle 1, 1 \rangle$, $\vec{y} = \langle -2, 1 \rangle$ **Answer:** $\text{dist}(\vec{x}, \vec{y}) = 3$
2. $\vec{x} = \langle 2, 3 \rangle$, $\vec{y} = \langle 0, 0 \rangle$ **Answer:** $\text{dist}(\vec{x}, \vec{y}) = \sqrt{13} \approx 3.61$
3. $\vec{x} = \left\langle 2, -\frac{1}{2} \right\rangle$, $\vec{y} = \langle 0, 8 \rangle$ **Answer:** $\text{dist}(\vec{x}, \vec{y}) = \frac{\sqrt{305}}{2} \approx 8.73$
4. $\vec{x} = \langle 1, -1 \rangle$, $\vec{y} = \langle -2, 2 \rangle$ **Answer:** $\text{dist}(\vec{x}, \vec{y}) = 3\sqrt{2} \approx 4.24$

Exercise 1.2.1:
Solution to Number 1: For the vectors $\vec{x} = \langle 1, 1, -1 \rangle$ and $\vec{y} = \langle -2, 1, 4 \rangle$, we have

$$2\vec{x} = \langle 2(1), 2(1), 2(-1) \rangle = \langle 2, 2, -2 \rangle$$

and

$$\vec{x} + \vec{y} = \langle 1, 1, -1 \rangle + \langle -2, 1, 4 \rangle = \langle 1 - 2, 1 + 1, -1 + 4 \rangle = \langle -1, 2, 3 \rangle$$

and

$$\begin{aligned}
\vec{x} - 3\vec{y} &= \langle 1, 1, -1 \rangle - 3\langle -2, 1, 4 \rangle \\
&= \langle 1, 1, -1 \rangle + \langle (-3)(-2), (-3)(1), (-3)(4) \rangle \\
&= \langle 1, 1, -1 \rangle + \langle 6, -3, -12 \rangle \\
&= \langle 7, -2, -13 \rangle .
\end{aligned}$$

Exercise 1.2.3: We want to prove that if

$$\vec{x} = \langle x_1, x_2, x_3 \rangle \quad \text{and} \quad \vec{y} = \langle y_1, y_2, y_3 \rangle$$

are two vectors in R^3, then

$$\|\vec{x} + \vec{y}\|^2 = \|\vec{x} - \vec{y}\|^2 \quad \text{if and only if} \quad \vec{x} \cdot \vec{y} = 0.$$

To prove this we first compute

$$\begin{aligned} \|\vec{x} + \vec{y}\|^2 &= \|\langle x_1 + y_1, x_2 + y_2, x_3 + y_3 \rangle\|^2 \\ &= (x_1 + y_1)^2 + (x_2 + y_2)^2 + (x_3 + y_3)^2 \\ &= x_1^2 + 2x_1y_1 + y_1^2 + x_2^2 + 2x_2y_2 + y_2^2 + x_3^2 + 2x_3y_3 + y_3^2 \\ &= x_1^2 + y_1^2 + x_2^2 + y_2^2 + x_3^2 + y_3^2 + 2x_1y_1 + 2x_2y_2 + 2x_3y_3 \\ &= x_1^2 + y_1^2 + x_2^2 + y_2^2 + x_3^2 + y_3^2 + 2\left(x_1y_1 + x_2y_2 + x_3y_3\right). \end{aligned}$$

A similar computation shows that

$$\|\vec{x} - \vec{y}\|^2 = x_1^2 + y_1^2 + x_2^2 + y_2^2 + x_3^2 + y_3^2 - 2\left(x_1y_1 + x_2y_2 + x_3y_3\right).$$

As can be seen from the above two computations $\|\vec{x} + \vec{y}\|^2 = \|\vec{x} - \vec{y}\|^2$ is true if an only

$$2\left(x_1y_1 + x_2y_2 + x_3y_3\right) = -2\left(x_1y_1 + x_2y_2 + x_3y_3\right)$$

and this is true if and only if

$$4\left(x_1y_1 + x_2y_2 + x_3y_3\right) = 0$$

which is true if and only if

$$x_1y_1 + x_2y_2 + x_3y_3 = 0.$$

Exercise 1.2.5:

$$\begin{aligned} \vec{x} &= \|\vec{x}\|\, \vec{x}_U \\ &= \sqrt{2}\left\langle \cos\left(90^\circ\right), \cos\left(45^\circ\right), \cos\left(45^\circ\right) \right\rangle \\ &= \sqrt{2}\left\langle 0, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right\rangle \\ &= \langle 0, 1, 1 \rangle \end{aligned}$$

Exercise 1.2.7:
The magnitude of $\vec{x}$ is

$$\|\vec{x}\| = \sqrt{(-3)^2 + 0^2 + 4^2} = 5$$

and the direction cosines of $\vec{x}$ are

$$\frac{x_1}{\|\vec{x}\|} = \frac{-3}{5}$$
$$\frac{x_2}{\|\vec{x}\|} = 0$$
$$\frac{x_3}{\|\vec{x}\|} = \frac{4}{5}$$

Thus the direction vector of $\vec{x}$ is

$$\vec{x}_U = \left\langle -\frac{3}{5}, 0, \frac{4}{5} \right\rangle.$$

Since $\vec{y}$ has magnitude 3 and points in the opposite direction of $\vec{x}$, then

$$\vec{y} = -3\vec{x}_U = -3 \left\langle -\frac{3}{5}, 0, \frac{4}{5} \right\rangle = \left\langle \frac{9}{5}, 0, -\frac{12}{5} \right\rangle.$$

Exercise 1.2.8 Find the distance between each pair of vectors.

1. $\vec{x} = \langle -3, 4, -5 \rangle$, $\vec{y} = \langle 0, 0, 0 \rangle$ **Answer:** $\text{dist}(\vec{x}, \vec{y}) = 5\sqrt{2} \approx 7.07$
2. $\vec{x} = \langle 1, 0, 1 \rangle$, $\vec{y} = \langle 3, -2, 1 \rangle$ **Answer:** $\text{dist}(\vec{x}, \vec{y}) = 2\sqrt{2} \approx 2.83$
3. $\vec{x} = \langle 1, 0, 0 \rangle$, $\vec{y} = \langle 0, 0, 1 \rangle$ **Answer:** $\text{dist}(\vec{x}, \vec{y}) = \sqrt{2} \approx 1.41$
4. $\vec{x} = \langle 2, -4, 5 \rangle$, $\vec{y} = \langle 0, 3, 3 \rangle$ **Answer:** $\text{dist}(\vec{x}, \vec{y}) = \sqrt{57} \approx 7.55$

Exercise 1.2.9 Let $\vec{x} = \langle 1, 0, 1 \rangle$ and $\vec{y} = \langle y_1, 3, -2 \rangle$. Find all values of y_1 such that $\text{dist}(\vec{x}, \vec{y}) = 8$.
Answer: We can square both sides of the equation $\text{dist}(\vec{x}, \vec{y}) = 8$ to avoid working with the radical.

$$(\text{dist}(\vec{x}, \vec{y}))^2 = (1 - y_1)^2 + (0 - 3)^2 + (1 - (-2))^2 = 8^2.$$

$$(1-y_1)^2 = 64-9-9 = 46.$$

Taking the square roots, $1-y_1 = \pm\sqrt{46}$, gives two solutions

$$y_1 = 1+\sqrt{46} \approx 7.78, \quad \text{or} \quad y_1 = 1-\sqrt{46} \approx -5.78$$

Exercise 1.3.1:
Solution for Number 3:
The vectors $\vec{x} = \langle -3,0,4,1,2\rangle$ and $\vec{y} = \langle 4,2,2,0,2\rangle$ are vectors in R^5. Also

$$\vec{x}+\vec{y} = \langle -3+4, 0+2, 4+2, 1+0, 2+2\rangle = \langle 1,2,6,1,4\rangle$$

$$\vec{x}-\vec{y} = \langle -3-4, 0-2, 4-2, 1-0, 2-2\rangle = \langle -7,-2,2,1,0\rangle$$

and

$$\vec{x}\cdot\vec{y} = (-3)(4)+(0)(2)+(4)(2)+(1)(0)+(2)(2) = 0.$$

The fact that $\vec{x}\cdot\vec{y} = 0$ tells us that $\vec{x}$ and $\vec{y}$ are orthogonal to each other. Let us verify that $\|\vec{x}+\vec{y}\|^2 = \|\vec{x}-\vec{y}\|^2$ (as must be the case since $\vec{x}$ and $\vec{y}$ have been found to be orthogonal to each other). We have

$$\|\vec{x}+\vec{y}\|^2 = 1^2+2^2+6^2+1^2+4^2 = 58$$

and

$$\|\vec{x}-\vec{y}\|^2 = (-7)^2+(-2)^2+2^2+1^1+0^2 = 58.$$

From this last result, $\text{dist}(\vec{x},\vec{y}) = \sqrt{58}$.

Exercise 1.3.3:
1) If $\vec{u}$ is a single vector in R^n, then Span $\{\vec{u}\}$ just means the set of all vectors in R^n that are scalar multiples of $\vec{u}$.

So if $\vec{u}$ is the vector $\vec{u} = \langle 1,0,1\rangle$ in R^3, then Span $\{\vec{u}\}$ is the set of all vectors in R^3 that are scalar multiples of $\vec{u}$. A scalar multiple of $\vec{u}$ is a vector of the form

$$c\vec{u} = c\langle 1,0,1\rangle = \langle c,0,c\rangle.$$

The vector $\vec{v} = \langle 2,0,2\rangle$ has the form $\langle c,0,c\rangle$ (with $c=2$) and thus $\vec{v}$ is in Span $\{\vec{u}\}$.

The vector $\vec{y} = \langle 1,0,2\rangle$ does not have the form $\langle c,0,c\rangle$ so $\vec{y}$ is not in Span $\{\vec{u}\}$.

The vector $\vec{0}_3 = \langle 0, 0, 0\rangle$ has the form $\langle c, 0, c\rangle$ (with $c = 0$) and thus $\vec{0}_3$ is in Span $\{\vec{u}\}$.

2) If $\vec{x}_1$ and $\vec{x}_2$ are any two vectors in R^n, then Span $\{\vec{x}_1, \vec{x}_2\}$ denotes the set of all possible linear combinations of $\vec{x}_1$ and $\vec{x}_2$. Thus Span $\{\vec{x}_1, \vec{x}_2\}$ consists of all vectors in R^n that have the form

$$c_1\vec{x}_1 + c_2\vec{x}_2$$

(where c_1 and c_2 are scalars).

Since $\vec{0}_n$ can be written as $\vec{0}_n = 0\vec{x}_1 + 0\vec{x}_2$, then $\vec{0}_n$ is in Span $\{\vec{x}_1, \vec{x}_2\}$.

Chapter 1 Additional Exercises

1. For $\vec{x}$ in R^n and scalar c in R, use the definition of the magnitude to show that $\|c\vec{x}\| = |c|\|\vec{x}\|$.

   Proof: Let c be a scalar and let $\vec{x} = \langle x_1, x_2, \ldots, x_n\rangle$ be a vector in R^n. Then

$$\begin{aligned}\|c\vec{x}\| &= \|\langle cx_1, cx_2, \ldots, cx_n\rangle\| \\ &= \sqrt{(cx_1)^2 + (cx_2)^2 + \cdots + (cx_n)^2} \\ &= \sqrt{c^2x_1^2 + c^2x_2^2 + \cdots + c^2x_n^2} \\ &= \sqrt{c^2\left(x_1^2 + x_2^2 + \cdots + x_n^2\right)} \\ &= \sqrt{c^2}\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \\ &= |c|\,\|\vec{x}\|\end{aligned}$$

2. Consider the vector $\vec{x} = \langle 1, -1, 0, 3\rangle$ in R^4. Determine the value(s) of p such that the vector $\vec{y} = \langle p, 1, 2, p\rangle$ is orthogonal to $\vec{x}$.

   Answer: $p = \frac{1}{4}$

3. Let $\vec{x} = \langle -2, 0, 2, 4, 5\rangle$, and $\vec{z} = \langle 4, 6, -3, 2, 2\rangle$. Find a vector $\vec{y}$ in R^5 such that

$$\vec{x} + \vec{y} = \vec{z}$$

   Answer: $\vec{y} = \langle 6, 6, -5, -2, -3\rangle$

4. For each pair of vectors, determine whether they are parallel, orthogonal, or neither parallel nor orthogonal.

   (a) $\vec{x} = \langle 1, -1, 3\rangle$, $\vec{y} = \langle -2, 2, -6\rangle$
   Answer: parallel

   (b) $\vec{x} = \langle 0, 4, 0, -2\rangle$, $\vec{y} = \langle 1, 2, 3, 4\rangle$
   Answer: orthogonal

   (c) $\vec{x} = \langle 1, 1, 0, 1, 1\rangle$, $\vec{y} = \langle -2, 2, -2, 2, 2\rangle$
   Answer: neither

   (d) $\vec{x} = \langle 2, -2, 8, 6, 12, 0\rangle$, $\vec{y} = \langle -1, 1, -4, -3, -6, 0\rangle$
   Answer: parallel

   (e) $\vec{x} = \langle 2, 0, -2, 1\rangle$, $\vec{y} = \langle 0, 1, 0, 0\rangle$
   Answer: orthogonal

5. Let $\vec{x} = \langle 1, 1, 2, 1\rangle$. Find all possible scalars, c such that $\|c\vec{x}\| = 1$.

   Answer: $c = \pm 1/\sqrt{7}$ which can also be written as $c = \pm\sqrt{7}/7$

6. Suppose that the vector $\vec{u}$ in R^n is orthogonal to every other vector in R^n. Explain why it must be that $\vec{u} = \langle 0, 0, \ldots, 0\rangle$. That is, $\vec{u} = \vec{0}_n$, the zero vector in R^n.

   Solution: Suppose $\vec{u} \neq \vec{0}$. Then at least one entry, say $u_i \neq 0$. Let $\vec{e}_i$ be the vector in R^n having i^{th} entry 1 and all other entries zero. Then the dot product $\vec{u} \cdot \vec{e}_i = u_i \neq 0$. But this contradicts our hypothesis that $\vec{u}$ is orthogonal to every other vector in R^n.

7. Let $\vec{u} = \langle -3, 5, 2\rangle$ and $\vec{x} = \langle 1, -1, -4\rangle$. Determine whether $\vec{y} = \langle 0, 1, -5\rangle$ is a linear combination of $\vec{u}$ and $\vec{x}$.

   Solution: $\vec{y}$ is a linear combination of $\vec{u}$ and $\vec{x}$ if and only if there exist scalars c and d such that $c\vec{u} + d\vec{x} = \vec{y}$. Thus let us look at the equation

   $$c\langle -3, 5, 2\rangle + d\langle 1, -1, -4\rangle = \langle 0, 1, -5\rangle$$

   which can be written as

   $$\langle -3c + d, 5c - d, 2c - 4d\rangle = \langle 0, 1, -5\rangle.$$

In order for the above equation to be satisfied, c and d must satisfy all three of the equations

$$-3c + d = 0 \tag{A.1}$$
$$5c - d = 1 \tag{A.2}$$
$$2c - 4d = -5. \tag{A.3}$$

Equation (A.1) requires that $d = 3c$ and when we substitute that into equation (A.2) we obtain

$$5c - 3c = 1$$

which gives $c = 1/2$. Since $d = 3c$, then we must have $d = 3/2$. We still need to make sure that equation (A.3) is satisfied by $(c, d) = (1/2, 3/2)$. It is, because

$$2\left(\frac{1}{2}\right) - 4\left(\frac{3}{2}\right) = -5.$$

We now see that $\vec{y}$ is a linear combination of $\vec{u}$ and $\vec{x}$ because

$$\vec{y} = \frac{1}{2}\vec{u} + \frac{3}{2}\vec{x}.$$

8. Let $\vec{z}_1 = \langle 1, 2\rangle$ and $\vec{z}_2 = \langle 2, 1\rangle$. Show that if $\vec{x} = \langle x_1, x_2\rangle$ is any vector in R^2, then $\vec{x}$ is in Span$\{\vec{z}_1, \vec{z}_2\}$. (Hint: find coefficients c_1 and c_2 such that $\vec{x} = c_1\vec{z}_1 + c_2\vec{z}_2$.)

9. For each statement, indicate whether the statement is true or false. Give a brief explanation or reason for each conclusion.

   (a) If $\vec{x}$ is a vector in R^4 such that $\|\vec{x}\| = 1$, then $\|2\vec{x}\| = 2^4$.
   Answer: This statement is false. If $\|\vec{x}\| = 1$, then

   $$\|2\vec{x}\| = |2|\,\|\vec{x}\| = 2\|\vec{x}\| = (2)\,(1) = 2.$$

   (b) For a vector $\vec{x}$ in R^n, the vector $-\vec{x}$ is equal to the vector $(-1)\,\vec{x}$.

   (c) For any pair of vectors $\vec{x}$and $\vec{y}$ in R^3, $\|\vec{x} + \vec{y}\| = \|\vec{x}\| + \|\vec{y}\|$.

Answer: This statement is false. For example, suppose that $\vec{x} = \langle 1,0,0 \rangle$ and $\vec{y} = \langle 0,1,0 \rangle$. Then $\|\vec{x}\| = 1$ and $\|\vec{y}\| = 1$ which means that $\|\vec{x}\| + \|\vec{y}\| = 1 + 1 = 2$. However

$$\begin{aligned} \|\vec{x} + \vec{y}\| &= \| \langle 1,0,0 \rangle + \langle 0,1,0 \rangle \| \\ &= \| \langle 1,1,0 \rangle \| \\ &= \sqrt{1^1 + 1^2 + 0^2} \\ &= \sqrt{2}. \end{aligned}$$

(d) If a vector $\vec{x}$ in $R^n$is orthogonal to itself, it must be the zero vector.

(e) If $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$is any set of vectors in R^n, then $\vec{0}_n$ is an element of $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$.

Answer: This statement is false. For example, consider the set of vectors $\{\vec{u}_1, \vec{u}_2\} = \langle \langle 1,2 \rangle, \langle 3,7 \rangle \rangle$ in R^2. Clearly, $\vec{0}_2$ is not an element of this set.

(f) If $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$is any set of vectors in R^n, then $\vec{0}_n$is an element of $\text{Span}\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$.

10. Let $\vec{x}$ be any nonzero element of R^5. Explain the difference between the set $\{\vec{x}\}$ and the set $\text{Span}\{\vec{x}\}$.

11. Use the dot product and the fact that $\|\vec{x}\|^2 = \vec{x} \cdot \vec{x}$ to prove the Pythagorean Theorem. The Pythagorean Theorem states

if $\vec{x}$ and $\vec{y}$ are orthogonal, then $\|\vec{x} + \vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2$.

Proof: Suppose that $\vec{x}$ and $\vec{y}$ are orthogonal to each other. Then $\vec{x} \cdot \vec{y} = 0$. This gives

$$\begin{aligned} \|\vec{x} + \vec{y}\|^2 &= (\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y}) \\ &= \vec{x} \cdot (\vec{x} + \vec{y}) + \vec{y} \cdot (\vec{x} + \vec{y}) \\ &= \vec{x} \cdot \vec{x} + \vec{x} \cdot \vec{y} + \vec{y} \cdot \vec{x} + \vec{y} \cdot \vec{y} \\ &= \vec{x} \cdot \vec{x} + \vec{y} \cdot \vec{y} + 2\,(\vec{x} \cdot \vec{y}) \\ &= \|\vec{x}\|^2 + \|\vec{y}\|^2 + 2\,(0) \\ &= \|\vec{x}\|^2 + \|\vec{y}\|^2. \end{aligned}$$

A.2 Chapter 2 Exercises:

Exercise 2.1.1

1. The solution set of the system

$$\begin{array}{rcrcrcrcr} 2x_1 & + & 4x_2 & + & 2x_3 & + & 2x_4 & = & -4 \\ x_1 & + & 2x_2 & + & 2x_3 & + & 6x_4 & = & -5 \end{array}$$

has parametric description

$$\begin{array}{rcl} x_1 & = & 1-2s+4t \\ x_2 & = & s, \\ x_3 & = & -3-5t \\ x_4 & = & t, \end{array} \qquad s,t \in R$$

Convert this to vector parametric form.
Answer: $\vec{x} = \langle 1,0,-3,0\rangle + s\langle -2,1,0,0\rangle + t\langle 4,0,-5,1\rangle, \quad s,t \in R$

2. The solution set of the system

$$\begin{array}{rcrcrcrcrcr} 3x_1 & + & x_2 & - & 2x_3 & + & 4x_4 & + & 2x_5 & = & -2 \\ x_1 & + & x_2 & + & 2x_3 & - & 2x_4 & + & x_5 & = & -4 \\ 2x_1 & - & x_2 & - & 8x_3 & + & 11x_4 & + & 2x_5 & = & -2 \end{array}$$

is the set of all five-tuples $(x_1, x_2, x_3, x_4, x_5)$ such that

$$x_1 = 4 + 2x_3 - 3x_4, \quad x_2 = -2 - 4x_3 + 5x_4, \quad x_5 = -6$$

and x_3 and x_4 can be any real number. Give a parametric description and a vector parametric description of the solution set.
Answer: A parametric description is

$$\begin{aligned} x_1 &= 4 + 2s - 3t \\ x_2 &= -2 - 4s + 5t \\ x_3 &= s \\ x_4 &= t \\ x_5 &= -6, \quad s,t \in R \end{aligned}$$

A vector parametric description is

$$\vec{x} = \langle 4,-2,0,0,-6\rangle + s\langle 2,-4,1,0,0\rangle + t\langle -3,5,0,1,0\rangle, \quad s,t \in R$$

Exercise 2.1.2 For each system, plot the lines determined by the equations together on the same set of axes and determine whether the system is inconsistent or consistent. If the system is consistent, state whether there is a unique solution or infinitely many solutions.

1. $\begin{array}{rcrcr} 3x_1 & + & x_2 & = & 0 \\ x_1 & - & 3x_2 & = & -1 \end{array}$

   Answer:

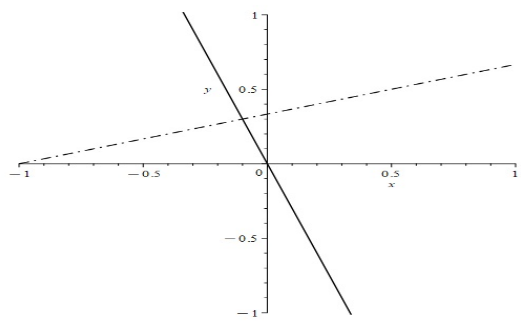


Figure A.1: Solid: $3x_1 + x_2 = 0$, Dash-dot $x_1 - 3x_2 = -1$

This system is consistent with a unique solution. The solution is the intersection $(x_1, x_2) = \left(-\frac{1}{10}, \frac{3}{10}\right)$.

3. $\begin{array}{rcrcr} 4x_1 & + & 6x_2 & = & 3 \\ 6x_1 & + & 9x_2 & = & 0 \end{array}$

   Answer:

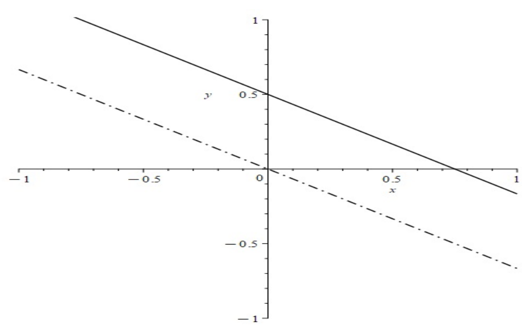


Figure A.2: Solid: $4x_1 + 6x_2 = 3$, Dash-dot $6x_1 + 9x_2 = 0$

This system is inconsistent. The lines are parallel.

4. $$\begin{array}{rcrcl} 6x_1 & + & 9x_2 & = & 0 \\ 4x_1 & + & 6x_2 & = & 0 \end{array}$$

Answer:

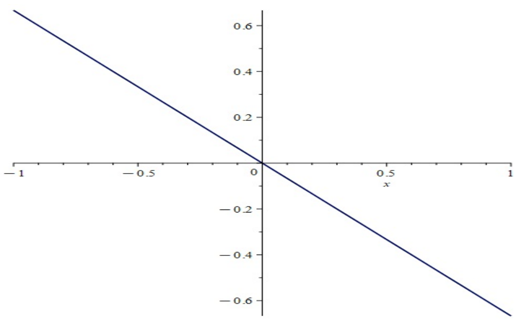


Figure A.3: Concurrent Lines $6x_1 + 9x_2 = 0$ and $4x_1 + 6x_3 = 0$

The system is consistent with infinitely many solutions. The solutions are all points on the common line, $\left\{(x_1, x_2) \,|\, x_1 = -\frac{3}{2}x_2,\ x_2 \in R\right\}$.

Exercise 2.1.3 Consider the system of two equations,

$$\begin{array}{rcrcl} a_{11}x_1 & + & a_{12}x_2 & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & = & b_2 \end{array}. \tag{A.4}$$

Explain why the system is guaranteed to be consistent with a unique solution whenever $a_{11}a_{22} \neq a_{21}a_{12}$.(Hint: A pair of lines in the plane are guaranteed to intersect exactly once if they have different slopes.)

Answer: The simplest case is if $a_{12}a_{22} \neq 0$ (meaning neither of the coefficients of x_2 is zero. In this case, the slopes are $m_1 = -\frac{a_{11}}{a_{12}}$ and $m_2 = -\frac{a_{21}}{a_{22}}$. Distinct slopes, $m_1 \neq m_2$, is equivalent to $a_{11}a_{22} \neq a_{21}a_{12}$. Suppose $a_{12} = 0$. Since $a_{11}a_{22} \neq a_{21}a_{12}$, we know that neither a_{11} nor a_{22} is zero. The first line is vertical, and the second line is not vertical. Hence the lines have different slopes and intersect exactly once. The case $a_{22} = 0$ is analogous with the second line being vertical.

Exercise 2.2.1 Perform the Gaussian elimination process on each system of equations. At each step, use the operation notation ($E_i \leftrightarrow E_j$, $kE_i \to E_i$, $kE_i + E_j \to E_j$) to clearly indicate the operation you have selected. If the system is consistent, state the solution in either parametric form or in vector parametric form.

Solution for problem 2.

2. $$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 2x_3 & = & 1 \\ 3x_1 & + & x_2 & - & x_3 & = & -2 \\ x_1 & + & x_2 & - & 2x_3 & = & 0 \end{array}$$

Answer:

$$-3E_1 + E_2 \to E_2 \qquad \begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 2x_3 & = & 1 \\ & - & 5x_2 & - & 7x_3 & = & -5 \\ x_1 & + & x_2 & - & 2x_3 & = & 0 \end{array}$$

$$-E_1 + E_3 \to E_3 \qquad \begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 2x_3 & = & 1 \\ & - & 5x_2 & - & 7x_3 & = & -5 \\ & - & x_2 & - & 4x_3 & = & -1 \end{array}$$

$$E_2 \leftrightarrow E_3 \qquad \begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 2x_3 & = & 1 \\ & - & x_2 & - & 4x_3 & = & -1 \\ & - & 5x_2 & - & 7x_3 & = & -5 \end{array}$$

$$-E_2 \to E_2 \qquad \begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 2x_3 & = & 1 \\ & & x_2 & + & 4x_3 & = & 1 \\ & - & 5x_2 & - & 7x_3 & = & -5 \end{array}$$

$$5E_2 + E_3 \to E_3 \qquad \begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 2x_3 & = & 1 \\ & & x_2 & + & 4x_3 & = & 1 \\ & & & & 13x_3 & = & 0 \end{array}$$

$$\tfrac{1}{13}E_3 \to E_3 \qquad \begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 2x_3 & = & 1 \\ & & x_2 & + & 4x_3 & = & 1 \\ & & & & x_3 & = & 0 \end{array}$$

The system is consistent. Performing the back substitution starting with $x_3 = 0$,

$$x_2 = 1 - 4x_3 = 1, \quad \text{and} \quad x_1 = 1 - 2x_2 - 2x_3 = -1.$$

In parametric form, we can write $\begin{array}{rcr} x_1 & = & -1 \\ x_2 & = & 1 \\ x_3 & = & 0 \end{array}$. In vector parametric form, the solution $\vec{x} = \langle -1, 1, 0 \rangle$.

Exercise 2.3.2 For each matrix A, write a homogeneous system of equations having A as its coefficient matrix.

1. $A = \begin{bmatrix} 1 & 0 & -1 & 2 \\ 2 & -3 & 2 & -1 \\ 0 & 2 & 4 & 2 \end{bmatrix}$
   Answer:
$$\begin{array}{rcrcrcrcl} x_1 & & & - & x_3 & + & 2x_4 & = & 0 \\ 2x_1 & - & 3x_2 & + & 2x_3 & - & 1x_4 & = & 0 \\ & & 2x_2 & + & 4x_3 & + & 2x_4 & = & 0 \end{array}$$

Exercise 2.3.3 For each augmented matrix A, write the corresponding system of equations.

2. $A = \left[\begin{array}{cc|c} 1 & 3 & 5 \\ 7 & 9 & 1 \\ 2 & 4 & 6 \end{array}\right]$
   Answer:
$$\begin{array}{rcrcl} x_1 & + & 3x_2 & = & 5 \\ 7x_1 & + & 9x_2 & = & 1 \\ 2x_1 & + & 4x_2 & = & 6 \end{array}$$

Exercise 2.3.4 Classify each matrix as a row echelon form (ref), a reduced row echelon form (rref), or not an echelon form. Identify which property (or properties) is not satisfied if a matrix is not an echelon form (or is an ref but not an rref).

1. $\begin{bmatrix} 1 & -1 \\ 0 & 3 \end{bmatrix}$ **Answer:** This is an ref but not an rref. The leading entry in row 2 is not 1, and it is not the only nonzero entry in its column.

3. $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 3 \end{bmatrix}$ **Answer:** This is not an ref. The first nonzero entry in each row is in the first column.

4. $\begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ **Answer:** This is an rref.

Exercise 2.3.5 Use the notation from Example 2.3.2 where appropriate.

1. Write out all possible 2×2 reduced row echelon forms.
   Answer:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & \square \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

3. Write out all of the possible 2×3 reduced row echelon forms.
   Answer:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & \square & \square \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & \square \\ 0 & 1 & \square \end{bmatrix},$$

$$\begin{bmatrix} 1 & \square & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & \square \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

5. Write out all possible 3×3 reduced row echelon forms.
   Answer:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & \square & \square \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & \square \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & \square \\ 0 & 1 & \square \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & \square & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Exercise 2.3.6 For each matrix A, follow the process outlined in the row reduction example to find rref(A).
Answer: (Detailed steps are given for problems 1. and 3.)

1. $A = \begin{bmatrix} -1 & 3 & 2 \\ 2 & -2 & 5 \end{bmatrix}$, $\text{rref}(A) = \begin{bmatrix} 1 & 0 & 19/4 \\ 0 & 1 & 9/4 \end{bmatrix}$

   Details (possible steps)

$$-R_1 \to R_1 \quad \begin{bmatrix} 1 & -3 & -2 \\ 2 & -2 & 5 \end{bmatrix}$$

$$-2R_1 + R_2 \to R_2 \quad \begin{bmatrix} 1 & -3 & -2 \\ 0 & 4 & 9 \end{bmatrix}$$

$$\tfrac{1}{4}R_2 \to R_2 \quad \begin{bmatrix} 1 & -3 & -2 \\ 0 & 1 & 9/4 \end{bmatrix}$$

$$3R_2 + R_1 \to R_1 \quad \begin{bmatrix} 1 & 0 & 19/4 \\ 0 & 1 & 9/4 \end{bmatrix}$$

2. $A = \begin{bmatrix} 4 & 4 & 0 & -2 \\ -1 & 3 & -5 & 1 \end{bmatrix}$, $\text{rref}(A) = \begin{bmatrix} 1 & 0 & 5/4 & -5/8 \\ 0 & 1 & -5/4 & 1/8 \end{bmatrix}$

3. $A = \begin{bmatrix} 2 & 4 & 6 & 8 \\ 4 & 6 & 8 & 10 \\ 6 & 8 & 10 & 4 \end{bmatrix}$, $\text{rref}(A) = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

   See Details on Next Page

4. $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$, $\text{rref}(A) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$

5. $A = \begin{bmatrix} 1 & 0 & -3 & 0 & 0 \\ 1 & 8 & -5 & -2 & 0 \\ 1 & 6 & -6 & 0 & -1 \\ 3 & 7 & -7 & -1 & -2 \end{bmatrix}$, $\text{rref}(A) = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1/3 \\ 0 & 0 & 1 & 0 & -1/3 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$

Detailed Solution for 3. (possible steps)

$$\begin{array}{l} \frac{1}{2}R_1 \to R_1 \\ \frac{1}{2}R_2 \to R_2 \\ \frac{1}{2}R_3 \to R_3 \end{array} \qquad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 2 \end{bmatrix}$$

$$\begin{array}{l} -2R_1 + R_2 \to R_2 \\ -3R_1 + R_3 \to R_3 \end{array} \qquad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -1 & -2 & -3 \\ 0 & -2 & -4 & -10 \end{bmatrix}$$

$$\begin{array}{l} -R_2 \to R_2 \\ -R_3 \to R_3 \end{array} \qquad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & 4 & 10 \end{bmatrix}$$

$$-2R_2 + R_3 \to R_3 \qquad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

$$\frac{1}{4}R_3 \to R_3 \qquad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{array}{l} -3R_3 + R_2 \to R_2 \\ -4R_3 + R_1 \to R_1 \end{array} \qquad \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$-2R_2 + R_1 \to R_1 \qquad \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Exercise 2.3.8 Suppose A is a 5×7 matrix.

1. If A is the coefficient matrix of a linear system of equations, how many variables does the system have?
   Answer: Seven. Each column holds the coefficients for one variable.

2. If A is the augmented matrix of a linear system of equations, how many variables does the system have?
   Answer: Six, the seventh column would be the constant terms.

3. Could A have 7 pivot columns? (Explain your answer.)
Answer: No. Each pivot position occupies a row as well as a column. With only five rows, it is not possible for there to be seven pivot positions (hence pivot columns).

Exercise 2.3.9 If A is an $m \times n$ matrix, what is the maximum number of pivot columns A can have? (Hint: consider the possible cases, $m < n$ and $m \geq n$. Explain your answer.)
Answer: Each pivot position occupies a row and column. The maximum number of pivot columns is the smaller of m and n, i.e., $\min\{m, n\}$.

Exercise 2.4.1 For each system of equations, use an augmented matrix and row reduction to either find the solutions set or determine that the system is inconsistent.

1. $$\begin{array}{rcrcrcrcr} x_1 & & & + & x_3 & & & = & 20 \\ & & x_2 & - & x_3 & - & x_4 & = & 0 \\ x_1 & + & x_2 & & & & & = & 80 \end{array}$$

$$\textbf{Answer:} \quad \begin{array}{rcl} x_1 & = & 20 - t \\ x_2 & = & 60 + t \\ x_3 & = & t \\ x_4 & = & 60 \end{array}, \quad -\infty < t < \infty$$

2. $$\begin{array}{rcrcrcl} x_1 & + & 2x_2 & + & 4x_3 & = & 0 \\ 2x_1 & + & 3x_2 & + & 5x_3 & = & 0 \\ 3x_1 & + & 4x_2 & + & 2x_3 & = & 0 \end{array} \qquad \textbf{Answer:} \quad \begin{array}{rcl} x_1 & = & 0 \\ x_2 & = & 0 \\ x_3 & = & 0 \end{array}$$

3. $$\begin{array}{rcrcrcr} 2x_1 & - & 2x_2 & + & x_3 & = & 6 \\ x_1 & + & x_2 & - & x_3 & = & -2 \\ & & x_2 & + & 3x_3 & = & 5 \end{array} \qquad \textbf{Answer:} \quad \begin{array}{rcl} x_1 & = & 1 \\ x_2 & = & -1 \\ x_3 & = & 2 \end{array}$$

Exercise 2.4.3 For each of the consistent systems in Exercise 2.4.2, write the solution set in parametric form. Either assign parameters to any free variables, or be sure to clearly indicate which variables (if any) are free.(Note: you may need to perform additional row operations.)

1. $$\begin{array}{rcl} x_1 & = & 1 - 2t \\ x_2 & = & t \\ x_3 & = & 2 \end{array}, \quad t \in R$$

2. $\begin{array}{rcl} x_1 &=& 1+t \\ x_2 &=& 2-3t \\ x_3 &=& t \\ x_4 &=& -4 \end{array}, \quad t \in R$

3. This system is inconsistent.

4. $\begin{array}{rcl} x_1 &=& -11/3 \\ x_2 &=& 10/3 \\ x_3 &=& -2/3 \end{array}$

5. $\begin{array}{rcl} x_1 &=& 4-2t+2s \\ x_2 &=& t \\ x_3 &=& 2-4s \\ x_4 &=& -2+5s \\ x_5 &=& -s \\ x_6 &=& s \end{array}, \quad s,t \in R$

6. $\begin{array}{rcl} x_1 &=& 2-2t+2s-2u \\ x_2 &=& t \\ x_3 &=& s \\ x_4 &=& -3+5u \\ x_5 &=& u \end{array}, \quad s,t,u \in R$

7. This system is inconsistent.

8. $\begin{array}{rcl} x_1 &=& -8 \\ x_2 &=& 5 \\ x_3 &=& t \end{array}, \quad t \in R$

Chapter 2 Additional Exercises

1. Solve each linear system by using row reduction on the associated augmented matrix.
   Note: Solutions are presented here in various forms: as an ordered n-tuple, in vector parametric, or in parametric form.

   a. $\begin{array}{rcrcrcl} x_1 &+& 2x_2 &+& x_3 &=& 1 \\ 3x_1 &+& 5x_2 &+& 3x_3 &=& 4 \\ 2x_1 &+& x_2 &+& x_3 &=& 4 \end{array}$
   Answer: $(2,-1,1)$

b. $$\begin{array}{rcrcrcr} x_1 & & & - & x_3 & = & 2 \\ 2x_1 & + & x_2 & + & 2x_3 & = & -6 \\ 3x_1 & + & 2x_2 & + & 2x_3 & = & -5 \end{array}$$

Answer: $\vec{x} = \langle -1, 2, -3 \rangle$

c. $$\begin{array}{rcrcrcrcr} -2x_1 & + & 2x_2 & - & 3x_3 & - & 2x_4 & = & -8 \\ 3x_1 & - & 3x_2 & + & 3x_3 & + & x_4 & = & 10 \\ 2x_1 & - & 2x_2 & + & 2x_3 & & & = & 4 \end{array}$$

Answer: $\begin{array}{rcl} x_1 & = & 6+t \\ x_2 & = & t \\ x_3 & = & -4 \\ x_4 & = & 4 \end{array}, \quad t \in R$

d. $$\begin{array}{rcrcrcrcrcr} -2x_1 & - & 6x_2 & + & 4x_3 & - & 8x_4 & + & 32x_5 & = & 18 \\ 3x_1 & + & 9x_2 & + & x_3 & - & 2x_4 & - & 6x_5 & = & 8 \end{array}$$

Answer:

$$\vec{x} = \langle 1, 0, 5, 0, 0, \rangle + s\langle -3, 1, 0, 0, 0 \rangle + t\langle 0, 0, 2, 1, 0 \rangle + \langle 4, 0, -6, 0, 1 \rangle$$

2. Determine all values of b, if any, such that the system of equations having the given augmented matrix is consistent.

   a. $\left[\begin{array}{rr|r} 2 & b & 3 \\ -1 & 3 & 4 \end{array}\right]$ **Answer** $b \neq -6$

   b. $\left[\begin{array}{rr|r} 4 & 3 & -2 \\ 6 & 1 & b \end{array}\right]$ **Answer** b is any real number

   c. $\left[\begin{array}{rr|r} 4 & 6 & b \\ 6 & 9 & 12 \end{array}\right]$ **Answer** $b = 8$

3. For each system of equations, determine all value(s) of b and c, if any, such that the system of equations has (i) no solution, (ii) a unique solution, and (iii) infinitely many solutions.

   a. $$\begin{array}{rcrcr} x_1 & + & 3x_2 & = & 2 \\ 3x_1 & + & bx_2 & = & c \end{array}$$

   Answer

   (i) $b = 9$ and $c \neq 6$,

   (ii) $b \neq 9$,

(iii) $b = 9$ and $c = 6$

b. $\begin{array}{rcrcl} bx_1 & - & 2x_2 & = & 5 \\ 4x_1 & + & 7x_2 & = & c \end{array}$

Answer

(i) $b = -8/7$ and $c \neq -35/2$,

(ii) $b \neq -8/7$,

(iii) $b = -8/7$ and $c = -35/2$

c. $\begin{array}{rcrcl} 3x_1 & + & bx_2 & = & 0 \\ cx_1 & + & 4x_2 & = & 0 \end{array}$

Answer

(i) the system consistent for all b and c,

(ii) all b and c such that $bc \neq 12$,

(iii) all b and c such that $bc = 12$

4. Create your own specific example of

   a. a system of linear equations with three equations and two variables that has a unique solution.
   Answer: Answers will vary.

   b. a system of linear equations with three equations and two variables that is inconsistent.
   Answer: Answers will vary.

   c. a system of linear equations with three equations and two variables that has infinitely many solutions.
   Answer: Answers will vary.

   d. a linear equation with one variable that has a unique solution.
   Answer: Answers will be some version of $ax_1 = b$ with $a \neq 0$.

   e. a linear equation with one variable that is inconsistent.
   Answer: Answers will be some version of the equation $0x_1 = b$ with $b \neq 0$.

   f. a linear equation with one variable that has infinitely many solutions.
   Answer: $0x_1 = 0$

5. Corollary 2.4.1 tells us that a system of linear equations that has more variables than equations either has no solution or has infinitely many

solutions. (Such a system cannot have a unique solution.) Create your own specific example of

a. a linear equation with two variables that has no solution.
Answer: Answers will be some version of $0x_1 + 0x_2 = b$ with $b \neq 0$

b. a linear equation with two variables that has infinitely many solutions.
Answer: Answers will vary.

c. a system of two linear equations with three variables that has no solution.
Answer: Answers will vary.

d. a system of two linear equations with three variables that has infinitely many solutions.
Answer: Answers will vary.

6. Find the solution set of the homogeneous system of linear equations having the given coefficient matrix.

a. $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ **Answer** $\begin{array}{rcl} x_1 & = & x_3 \\ x_2 & = & -2x_3 \\ x_3 & \text{is} & \text{free} \end{array}$

b. $\begin{bmatrix} 2 & -1 & 7 \\ 3 & 1 & 3 \end{bmatrix}$ **Answer** $\begin{array}{rcl} x_1 & = & -2t \\ x_2 & = & 3t \\ x_3 & = & t \end{array}, \quad t \in R$

c. $\begin{bmatrix} 1 & 2 & 1 \\ 3 & 5 & 3 \\ 2 & 1 & 1 \end{bmatrix}$ **Answer** $(0, 0, 0)$

d. $\begin{bmatrix} 3 & 9 & 1 & -2 \\ 1 & 3 & -2 & 4 \end{bmatrix}$ **Answer** $\begin{array}{rcl} x_1 & = & -3t \\ x_2 & = & t \\ x_3 & = & 2s \\ x_4 & = & s \end{array}, \quad t, s \in R$

e. $\begin{bmatrix} 1 & 3 & 4 \\ -1 & -5 & -7 \\ 2 & 4 & 5 \\ 3 & 3 & 3 \end{bmatrix}$ **Answer** $\begin{array}{rcl} x_1 & = & \frac{1}{2}x_3 \\ x_2 & = & -\frac{3}{2}x_3 \\ x_3 & \text{is} & \text{free} \end{array}$

7. Find an equation satisfied by g, h, and k such that the given matrix is the augmented matrix of a consistent linear system

$$\left[\begin{array}{rrr|c} 1 & -4 & 7 & g \\ 0 & 3 & -5 & h \\ -2 & 5 & -9 & k \end{array}\right]$$

Answer Row reduction to an ref results in a row of the form

$$\left[0\ 0\ 0 \mid 2g + h + k\right].$$

For the system to be consistent, the entry in the augment column must be zero. So an equation in the parameters g, h and k is $2g + h + k = 0$.

8. Propane combines with oxygen to form carbon dioxide and water according to the chemical equation

$$x_1\ C_3H_8 + x_2\ O_2 \quad \longrightarrow \quad x_3\ CO_2 + x_4\ H_2O.$$

Balancing the number of atoms of carbon (C), hydrogen (H), and oxygen (O) leads to the homogeneous system of equations

$$\begin{array}{rcrcr} 3x_1 & = & x_3 & & \\ 8x_1 & = & 2x_4 & & \\ 2x_2 & = & 2x_3 & + & x_4 \end{array} .$$

Show that this system is homogeneous. Find the smallest positive integers x_1, x_2, x_3, x_4 that balance the chemical equation.
Answer If we write the system in the standard format, having all variables on one side, the system is

$$\begin{array}{rcrcrcrcr} 3x_1 & & & - & x_3 & & & = & 0 \\ 8x_1 & & & & & - & 2x_4 & = & 0 \\ & & 2x_2 & - & 2x_3 & - & x_4 & = & 0 \end{array} .$$

Every constant term is zero making the system homogeneous. The solution is given parametrically by

$$\begin{array}{rcl} x_1 & = & \frac{1}{4}t \\ x_2 & = & \frac{5}{4}t \\ x_3 & = & \frac{3}{4}t \\ x_4 & = & t \end{array}, \quad t \in R.$$

To obtain positive integers, we see that t must be a multiple of 4. The smallest set of coefficients is obtained by taking $t = 4$. Then $\vec{x} = \langle 1, 5, 3, 4 \rangle$. The balanced equation is

$$C_3H_8 + 5\ O_2 \quad \longrightarrow \quad 3\ CO_2 + 4\ H_2O.$$

9. Boron sulfide and water react to produce boric acid and hydrogen sulfide gas according to the chemical equation

$$x_1\ B_2S_3 + x_2\ H_2O \quad \longrightarrow \quad x_3\ H_3BO_3 + x_4\ H_2S.$$

Balancing the number of atoms of boron (B), sulfer (S), hydrogen (H) and oxygen (O) leads to the homogeneous system of equations

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

Show that this system is homogeneous. Find the smallest positive integers x_1, x_2, x_3, x_4 that balance the chemical equation.

Answer If we arrange the equations with all variables on the left, we see that all constant terms are zero making the system homogeneous.

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

The solution of this system is given parametrically by

$$\begin{array}{rcl} x_1 & = & \frac{1}{3}t \\ x_2 & = & 2t \\ x_3 & = & \frac{2}{3}t \\ x_4 & = & t \end{array}, \quad t \in R.$$

The smallest positive integer solution requires $t = 3$ giving $\vec{x} = \langle 1, 6, 2, 3 \rangle$. The balanced chemical equation is

$$B_2S_3 + 6\ H_2O \quad \longrightarrow \quad 2\ H_3BO_3 + 3\ H_2S.$$

$$\begin{aligned} x_1 &= 1 \\ x_2 &= 6 \\ x_3 &= 2 \\ x_4 &= 3 \end{aligned}$$

10. Suppose A is an $m \times n$ matrix whose i^{th} column is all zeros. Explain why the i^{th} column of rref(A) is all zeros.

    Answer Any row operation performed on the matrix will result in some linear combination of zeros replacing each entry in column i. Since every linear combination of zeros is zero, the entries in that column will remain zero.

11. Let

    $$\vec{a}_1 = \langle 1, 0, 1, 0\rangle, \;\; \vec{a}_2 = \langle -1, 2, 1, 1\rangle, \;\; \vec{a}_3 = \langle 0, 0, 2, 2\rangle, \;\; \text{and} \;\; \vec{a}_4 = \langle 1, 1, 0, -1\rangle.$$

    Show that the vector $\vec{y} = \langle 2, -1, 3, 3\rangle$ in R^4 is a linear combination of the vectors $\vec{a}_1, \vec{a}_2, \vec{a}_3$ and $\vec{a}_4$, and identify the weights. (Hint: the equation $x_1\vec{a}_1 + x_1\vec{a}_1 + x_1\vec{a}_1 + x_1\vec{a}_1 = \vec{y}$ can be translated into a linear system of equations for the weights $x_1, \ldots, x_4$.)

    Answer We use the algebra defined for vectors in R^4 in Chapter 1. Note that

    $$x_1\vec{a}_1 + x_1\vec{a}_1 + x_1\vec{a}_1 + x_1\vec{a}_1 =$$

    $$x_1\langle 1, 0, 1, 0\rangle + x_2\langle -1, 2, 1, 1\rangle + x_3\langle 0, 0, 2, 2\rangle + x_4\langle 1, 1, 0, -1\rangle =$$

    $$\langle x_1 - x_2 + x_4, 2x_2 + x_4, x_1 + x_2 + 2x_3, x_2 + 2x_3 - x_4\rangle.$$

    Now, we set this linear combination equal to the vector $\vec{y}$ and create an equation for each of the four entries.

    $$\langle x_1 - x_2 + x_4, 2x_2 + x_4, x_1 + x_2 + 2x_3, x_2 + 2x_3 - x_4\rangle = \langle 2, -1, 3, 3\rangle$$

    which implies

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

We can use an augmented matrix with row reduction to show that the system is consistent and to find a solution.

$$\left[\begin{array}{rrrr|r} 1 & -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 1 & -1 \\ 1 & 1 & 2 & 0 & 3 \\ 0 & 1 & 2 & -1 & 3 \end{array}\right] \xrightarrow{rref} \left[\begin{array}{rrrr|r} 1 & 0 & 0 & 0 & -3 \\ 0 & 1 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 3 \end{array}\right]$$

Hence the system is consistent, meaning $\vec{y}$ can be written as a linear combination of $\vec{a}_1, \vec{a}_2, \vec{a}_3$ and $\vec{a}_4$ with weights $x_1 = -3$ $x_2 = -2$, $x_3 = 4$ and $x_4 = 3$. That is

$$-3\vec{a}_1 - 2\vec{a}_2 + 4\vec{a}_3 + 3\vec{a}_4 = \vec{y}.$$

12. Determine whether the vector $\vec{x} = \langle -1, 3, 1\rangle$ in R^3 is a linear combination of the vectors $\vec{u}$ and $\vec{v}$, where

$$\vec{u} = \langle 1, 1, -2\rangle, \text{ and } \vec{v} = \langle 3, 2, 2\rangle.$$

Answer: It is not. If we set up the equation

$$\vec{x} = c_1\vec{u} + c_2\vec{v},$$

we get a system of three equations,

$$\begin{array}{rcrcr} c_1 & + & 3c_2 & = & -1 \\ c_1 & + & 2c_2 & = & 3 \\ -2c_1 & + & 2c_2 & = & 1 \end{array}.$$

Setting up and reducing the augmented matrix,

$$\left[\begin{array}{rr|r} 1 & 3 & -1 \\ 1 & 2 & 3 \\ -2 & 2 & 1 \end{array}\right] \xrightarrow{rref} \left[\begin{array}{rr|r} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right].$$

The system is inconsistent, so $\vec{x}$ is not a linear combination of $\vec{u}$ and $\vec{v}$.

A.3 Chapter 3 Exercises:

Exercise 3.1.1

For the 5×4 matrix

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

the entry in row 3 and column 3 of A is 6. We can express this fact by writing either

$$a_{33} = 6 \quad \text{or} \quad A_{(3,3)} = 6.$$

Likewise, we can write either

$$a_{24} = 3 \quad \text{or} \quad A_{(2,4)} = 3.$$

Exercise 3.1.2:

1) The row vectors of the matrix

$$A = \begin{bmatrix} 8 & 4 \\ -5 & -5 \\ 3 & -5 \\ 8 & 5 \end{bmatrix}$$

are vectors in R^2. The column vectors of A are vectors in R^4.

Exercise 3.2.1

Solutions for numbers 3 and 4:

For the matrices that A and B are the matrices

$$A = \begin{bmatrix} -2 & 3 & -3 \\ 3 & 5 & 3 \\ 3 & 5 & -5 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 1 & -3 \\ 3 & -5 & 5 \end{bmatrix}$$

and the scalars $c = -2$ and $d = 2$, we have

$$\begin{aligned} B - A &= \begin{bmatrix} 2 & 0 & 1 \\ 1 & 1 & -3 \\ 3 & -5 & 5 \end{bmatrix} - \begin{bmatrix} -2 & 3 & -3 \\ 3 & 5 & 3 \\ 3 & 5 & -5 \end{bmatrix} \\ &= \begin{bmatrix} 2-(-2) & 0-3 & 1-(-3) \\ 1-3 & 1-5 & -3-3 \\ 3-3 & -5-5 & 5-(-5) \end{bmatrix} \\ &= \begin{bmatrix} 4 & -3 & 4 \\ -2 & -4 & -6 \\ 0 & -10 & 10 \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} cA &= -2A \\ &= -2\begin{bmatrix} -2 & 3 & -3 \\ 3 & 5 & 3 \\ 3 & 5 & -5 \end{bmatrix} \\ &= \begin{bmatrix} (-2)(-2) & (-2)(3) & (-2)(-3) \\ (-2)(3) & (-2)(5) & (-2)(3) \\ (-2)(3) & (-2)(5) & (-2)(-5) \end{bmatrix} \\ &= \begin{bmatrix} 4 & -6 & 6 \\ -6 & -10 & -6 \\ -6 & -10 & 10 \end{bmatrix}. \end{aligned}$$

Exercise 3.2.3:
Solution for number 1: For $c = -5$ and

$$A = \begin{bmatrix} 1 & 0 \\ -2 & -3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 \\ -3 & -3 \end{bmatrix},$$

we have

$$\begin{aligned} c(A+B) &= -5\left(\begin{bmatrix} 1 & 0 \\ -2 & -3 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ -3 & -3 \end{bmatrix}\right) \\ &= -5\begin{bmatrix} 2 & 1 \\ -5 & -6 \end{bmatrix} \\ &= \begin{bmatrix} -10 & -5 \\ 25 & 30 \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} cA + cB &= -5\begin{bmatrix} 1 & 0 \\ -2 & -3 \end{bmatrix} + (-5)\begin{bmatrix} 1 & 1 \\ -3 & -3 \end{bmatrix} \\ &= \begin{bmatrix} -5 & 0 \\ 10 & 15 \end{bmatrix} + \begin{bmatrix} -5 & -5 \\ 15 & 15 \end{bmatrix} \\ &= \begin{bmatrix} -10 & -5 \\ 25 & 30 \end{bmatrix}. \end{aligned}$$

We have verified that

$$c(A+B) = cA + cB = \begin{bmatrix} -10 & -5 \\ 25 & 30 \end{bmatrix}.$$

Exercise 3.3.1
Solutions for numbers 3 and 6:
3) For

$$A = \begin{bmatrix} 2 & 5 \\ 0 & 7 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & -4 & -7 \\ -7 & 0 & -1 \end{bmatrix}$$

we have

$$AB = \begin{bmatrix} 2 & 5 \\ 0 & 7 \end{bmatrix}\begin{bmatrix} 2 & -4 & -7 \\ -7 & 0 & -1 \end{bmatrix} = \begin{bmatrix} -31 & -8 & -19 \\ -49 & 0 & -7 \end{bmatrix}$$

and BA is not defined (because B has 3 columns and A has 2 rows).
6) For

$$A = \begin{bmatrix} 2 & 3 \\ -4 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 3 \\ -4 & 0 \end{bmatrix},$$

we have

$$AB = \begin{bmatrix} 2 & 3 \\ -4 & 2 \end{bmatrix}\begin{bmatrix} 0 & 3 \\ -4 & 0 \end{bmatrix} = \begin{bmatrix} -12 & 6 \\ -8 & -12 \end{bmatrix}$$

and

$$BA = \begin{bmatrix} 0 & 3 \\ -4 & 0 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ -4 & 2 \end{bmatrix} = \begin{bmatrix} -12 & 6 \\ -8 & -12 \end{bmatrix}.$$

We see that $AB = BA$.
Exercise 3.3.3
Solution for number 1: For the matrices

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

we have

$$\begin{aligned} A(B+C) &= \begin{bmatrix} 3 & 0 \\ 1 & 3 \end{bmatrix} \left(\begin{bmatrix} 1 & 5 \\ 2 & 0 \end{bmatrix} + \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix} \right) \\ &= \begin{bmatrix} 3 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 6 \\ 2 & -2 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 18 \\ 6 & 0 \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} AB + AC &= \begin{bmatrix} 3 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 5 \\ 2 & 0 \end{bmatrix} + \begin{bmatrix} 3 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 15 \\ 7 & 5 \end{bmatrix} + \begin{bmatrix} -3 & 3 \\ -1 & -5 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 18 \\ 6 & 0 \end{bmatrix}. \end{aligned}$$

We have verified that

$$A(B+C) = AB + AC = \begin{bmatrix} 0 & 18 \\ 6 & 0 \end{bmatrix}.$$

Exercise 3.3.4 We can verify that a given entry in $(A+B)C$ is the same as the corresponding entry in $AC+BC$. If we consider the entry in row i and column j of $(A+B)C$, we have

$$\big((A+B)C\big)_{(i,j)} = \mathrm{Row}_i(A+B) \cdot \mathrm{Col}_j(C).$$

But $\mathrm{Row}_i(A+B) = \mathrm{Row}_i(A) + \mathrm{Row}_i(B)$ from the definition of matrix addition. Using the given distributive property of the dot product,

$$\begin{aligned} \big((A+B)C\big)_{(i,j)} &= \mathrm{Row}_i(A+B) \cdot \mathrm{Col}_j(C) \\ &= \big(\mathrm{Row}_i(A) + \mathrm{Row}_i(B)\big) \cdot \mathrm{Col}_j(C) \\ &= \mathrm{Row}_i(A) \cdot \mathrm{Col}_j(C) + \mathrm{Row}_i(B) \cdot \mathrm{Col}_j(C) \end{aligned}$$

The first term in the sum is the entry $\big(AC\big)_{(i,j)}$ in the product AC, and the second term is the entry $\big(BC\big)_{(i,j)}$ in the product BC. So their sum is the entry in row i and column j of the sum $AC+BC$. That is

$$\big((A+B)C\big)_{(i,j)} = \big(AC+BC\big)_{(i,j)}.$$

This holds for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$, so we can conclude that $(A+B)C = AC + BC$.

Exercise 3.3.5 For the matrices

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

we have

$$\begin{aligned} (A+B)C &= \left(\begin{bmatrix} 3 & 0 \\ 1 & 3 \end{bmatrix} + \begin{bmatrix} 1 & 5 \\ 2 & 0 \end{bmatrix} \right) \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix} \\ &= \begin{bmatrix} 4 & 5 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix} \\ &= \begin{bmatrix} -4 & -6 \\ -3 & -3 \end{bmatrix}. \end{aligned}$$

Also,

$$\begin{aligned} AC + BC &= \begin{bmatrix} 3 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix} + \begin{bmatrix} 1 & 5 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix} \\ &= \begin{bmatrix} -3 & 3 \\ -1 & -5 \end{bmatrix} + \begin{bmatrix} -1 & -9 \\ -2 & 2 \end{bmatrix} \\ &= \begin{bmatrix} -4 & -6 \\ -3 & -3 \end{bmatrix}. \end{aligned}$$

This shows that

$$(A+B)C = AC + BC = \begin{bmatrix} -4 & -6 \\ -3 & -3 \end{bmatrix}.$$

Exercise 3.4.1:
If A has size 5×7 and B has size 7×3, then A^T has size 7×5, B^T has size 3×7, and $(AB)^T$ has size 3×5.
Exercise 3.4.3:
Solution for number 1:
For

$$A = \begin{bmatrix} -4 & 0 \\ 1 & 2 \\ -3 & 1 \end{bmatrix} \text{ and } c = 3,$$

we have

$$\begin{aligned} (cA)^T &= \left(3\begin{bmatrix} -4 & 0 \\ 1 & 2 \\ -3 & 1 \end{bmatrix}\right)^T \\ &= \begin{bmatrix} -12 & 0 \\ 3 & 6 \\ -9 & 3 \end{bmatrix}^T \\ &= \begin{bmatrix} -12 & 3 & -9 \\ 0 & 6 & 3 \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} c\left(A^T\right) &= 3\left(\begin{bmatrix} -4 & 0 \\ 1 & 2 \\ -3 & 1 \end{bmatrix}^T\right) \\ &= 3\begin{bmatrix} -4 & 1 & -3 \\ 0 & 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -12 & 3 & -9 \\ 0 & 6 & 3 \end{bmatrix}. \end{aligned}$$

We have shown that

$$(cA)^T = c\left(A^T\right) = \begin{bmatrix} -12 & 3 & -9 \\ 0 & 6 & 3 \end{bmatrix}.$$

Exercise 3.4.5
Answer to the First Question: Only the first row of the matrix

$$A = \begin{bmatrix} 1 & 2 & -6 \\ \text{--} & \text{--} & \text{--} \\ \text{--} & \text{--} & \text{--} \end{bmatrix},$$

is given. Is it possible to fill in the remaining two rows of A in such a way that the statement $A^T = A$ is true? If so, then do it. If not, then explain why not.

Answer: It can be done. One way to do it is to let

$$A = \begin{bmatrix} 1 & 2 & -6 \\ 2 & 4 & 8 \\ -6 & 8 & 53 \end{bmatrix}$$

and we have $A^T = A$.

Exercise 3.5.1:

Solutions for numbers 3, 7, and 9:

3) For

$$A = \begin{bmatrix} -2 & 2 & 5 & -1 \\ 4 & 0 & 2 & -1 \\ 1 & -1 & 2 & 1 \end{bmatrix}, \qquad \vec{x} = \langle 2, -3, -3, 5 \rangle ,$$

we can compute $A\vec{x}$ by computing

$$\begin{aligned} \mathrm{Row}_1(A) \cdot \vec{x} &= \langle -2, 2, 5, -1 \rangle \cdot \langle 2, -3, -3, 5 \rangle = -30 \\ \mathrm{Row}_2(A) \cdot \vec{x} &= \langle 4, 0, 2, -1 \rangle \cdot \langle 2, -3, -3, 5 \rangle = -3 \\ \mathrm{Row}_3(A) \cdot \vec{x} &= \langle 1, -1, 2, 1 \rangle \cdot \langle 2, -3, -3, 5 \rangle = 4 \end{aligned}$$

to obtain

$$A\vec{x} = \langle \mathrm{Row}_1(A) \cdot \vec{x}, \mathrm{Row}_2(A) \cdot \vec{x}, \mathrm{Row}_3(A) \cdot \vec{x} \rangle = \langle -30, -3, 4 \rangle .$$

We can also compute $A\vec{x}$ by computing

$$\begin{aligned} 2\,\mathrm{Col}_1(A) &= \langle -4, 8, 2 \rangle \\ -3\,\mathrm{Col}_2(A) &= \langle -6, 0, 3 \rangle \\ -3\,\mathrm{Col}_3(A) &= \langle -15, -6, -6 \rangle \\ 5\,\mathrm{Col}_4(A) &= \langle -5, -5, 5 \rangle \end{aligned}$$

to obtain

$$\begin{aligned} A\vec{x} &= x_1\,\mathrm{Col}_1(A) + x_2\,\mathrm{Col}_2(A) + x_3\,\mathrm{Col}_3(A) + x_4\,\mathrm{Col}_4(A) + x_5\,\mathrm{Col}_5(A) \\ &= \langle -4, 8, 2 \rangle + \langle -6, 0, 3 \rangle + \langle -15, -6, -6 \rangle + \langle -5, -5, 5 \rangle \\ &= \langle -30, -3, 4 \rangle \end{aligned}$$

7 and 9) A matrix A and the vector $\vec{x}$ such that $A\vec{x} =$

$$\langle \langle 6, -3, -2 \rangle \cdot \langle 1, -3, 1 \rangle , \langle 5, -2, 2 \rangle \cdot \langle 1, -3, 1 \rangle , \langle -3, -3, -1 \rangle \cdot \langle 1, -3, 1 \rangle \rangle$$

are

$$A = \begin{bmatrix} 6 & -3 & -2 \\ 5 & -2 & 2 \\ -3 & -3 & -1 \end{bmatrix} \quad \text{and} \quad \vec{x} = \langle 1, -3, 1 \rangle .$$

Here is how we write $A\vec{x}$ as a linear combination of the column vectors of A:

$$A\vec{x} = (1)\langle 6,5,-3\rangle + (-3)\langle -3,-2,-3\rangle + (1)\langle -2,2,-1\rangle .$$

Exercise 3.5.3:

Suppose that A is a 4×5 matrix and that B is a 5×3 matrix and that the second column of B consists entirely of entries of 0.

1. What size is the matrix AB?

   Answer: AB has size 4×3.

2. What can you say about the second column of AB?

   Answer: The second column of AB is

   $$\mathrm{Col}_2(AB) = A\,\mathrm{Col}_2(B).$$

   Since we are given that $\mathrm{Col}_2(B) = \vec{0}_5$, then

   $$\mathrm{Col}_2(AB) = A\vec{0}_5 = \vec{0}_4.$$

Exercise 3.5.4

We will give solutions for numbers 6, 8, and 10.

6) For

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix}, \qquad \vec{x} = \langle x_1, x_2\rangle ,$$

we can compute $A^T\vec{x}$ by computing

$$\begin{aligned}
\mathrm{Col}_1(A)\cdot\vec{x} &= \langle a_{11},a_{21}\rangle\cdot\langle x_1,x_2\rangle = a_{11}x_1 + a_{21}x_2\\
\mathrm{Col}_2(A)\cdot\vec{x} &= \langle a_{12},a_{22}\rangle\cdot\langle x_1,x_2\rangle = a_{12}x_1 + a_{22}x_2\\
\mathrm{Col}_3(A)\cdot\vec{x} &= \langle a_{13},a_{23}\rangle\cdot\langle x_1,x_2\rangle = a_{13}x_1 + a_{23}x_2\\
\mathrm{Col}_4(A)\cdot\vec{x} &= \langle a_{14},a_{24}\rangle\cdot\langle x_1,x_2\rangle = a_{14}x_1 + a_{24}x_2
\end{aligned}$$

to obtain

$$\begin{aligned}
A^T\vec{x} &= \langle \mathrm{Col}_1(A)\cdot\vec{x}, \mathrm{Col}_2(A)\cdot\vec{x}, \mathrm{Col}_3(A)\cdot\vec{x}, \mathrm{Col}_4(A)\cdot\vec{x}\rangle\\
&= \langle a_{11}x_1 + a_{21}x_2, a_{12}x_1 + a_{22}x_2, a_{13}x_1 + a_{23}x_2, a_{14}x_1 + a_{24}x_2\rangle .
\end{aligned}$$

We can also compute $A^T\vec{x}$ by computing

$$\begin{aligned} x_1 \operatorname{Row}_1(A) &= x_1 \langle a_{11}, a_{12}, a_{13}, a_{14}\rangle = \langle a_{11}x_1, a_{12}x_1, a_{13}x_1, a_{14}x_1\rangle \\ x_2 \operatorname{Row}_2(A) &= x_2 \langle a_{21}, a_{22}, a_{23}, a_{24}\rangle = \langle a_{21}x_2, a_{22}x_2, a_{23}x_2, a_{24}x_2\rangle \end{aligned}$$

to obtain

$$\begin{aligned} A^T\vec{x} &= x_1 \operatorname{Row}_1(A) + x_2 \operatorname{Row}_2(A) \\ &= \langle a_{11}x_1, a_{12}x_1, a_{13}x_1, a_{14}x_1\rangle + \langle a_{21}x_2, a_{22}x_2, a_{23}x_2, a_{24}x_2\rangle \\ &= \langle a_{11}x_1 + a_{21}x_2, a_{12}x_1 + a_{22}x_2, a_{13}x_1 + a_{23}x_2, a_{14}x_1 + a_{24}x_2\rangle . \end{aligned}$$

8 and 10) A matrix A and the vector $\vec{x}$ such that

$$A^T\vec{x} = (-3)\langle -4, -1\rangle + (-7)\langle 1, 0\rangle + (2)\langle 2, -2\rangle$$

are

$$A = \begin{bmatrix} -4 & -1 \\ 1 & 0 \\ 2 & -2 \end{bmatrix} \quad \text{and} \quad \vec{x} = \langle -3, -7, 2\rangle .$$

We can also write $A^T\vec{x}$ as follows:

$$A^T\vec{x} = \langle \langle -4, 1, 2\rangle \cdot \langle -3, -7, 2\rangle , \langle -1, 0, -2\rangle \cdot \langle -3, -7, 2\rangle\rangle$$

Exercise 3.5.5
For the matrices

$$A = \begin{bmatrix} 1 & -4 \\ 1 & 4 \end{bmatrix} \text{ and } B = \begin{bmatrix} -4 & 2 \\ -2 & -4 \end{bmatrix},$$

we want to verify by computation that

$$\operatorname{Row}_1(AB) = B^T \operatorname{Row}_1(A)$$

and

$$\operatorname{Row}_2(AB) = B^T \operatorname{Row}_2(A).$$

So let us do the computations.
We have

$$AB = \begin{bmatrix} 1 & -4 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} -4 & 2 \\ -2 & -4 \end{bmatrix} = \begin{bmatrix} 4 & 18 \\ -12 & -14 \end{bmatrix}$$

from which we see that

$$\begin{aligned}\text{Row}_1(AB) &= \langle 4, 18\rangle \\ \text{Row}_2(AB) &= \langle -12, -14\rangle .\end{aligned}$$

Also

$$\begin{aligned}B^T\,\text{Row}_1(A) &= \begin{bmatrix} -4 & -2 \\ 2 & -4 \end{bmatrix}\langle 1, -4\rangle \\ &= (1)\langle -4, 2\rangle + (-4)\langle -2, -4\rangle \\ &= \langle -4, 2\rangle + \langle 8, 16\rangle \\ &= \langle 4, 18\rangle\end{aligned}$$

and

$$\begin{aligned}B^T\,\text{Row}_2(A) &= \begin{bmatrix} -4 & -2 \\ 2 & -4 \end{bmatrix}\langle 1, 4\rangle \\ &= (1)\langle -4, 2\rangle + (4)\langle -2, -4\rangle \\ &= \langle -4, 2\rangle + \langle -8, -16\rangle \\ &= \langle -12, -14\rangle\end{aligned}$$

Exercise 3.6.1:

1) Write down the four standard unit vectors in R^4.

   Answer: The four standard unit vectors in R^4 are

$$\begin{aligned}\vec{e}_1 &= \langle 1, 0, 0, 0\rangle \\ \vec{e}_2 &= \langle 0, 1, 0, 0\rangle \\ \vec{e}_3 &= \langle 0, 0, 1, 0\rangle \\ \vec{e}_4 &= \langle 0, 0, 0, 1\rangle .\end{aligned}$$

3) Let $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ be the set of standard unit vectors in R^3 and let A be the matrix

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

Compute $A\vec{e}_1$, $A\vec{e}_2$,and $A\vec{e}_3$. Write out the computations in detail. You should observe that $A\vec{e}_i = \mathrm{Col}_i(A)$ for all $i = 1, 2, 3$.

Solution:

$$\begin{aligned}
A\vec{e}_1 &= (1)\,\mathrm{Col}_1(A) + (0)\,\mathrm{Col}_2(A) + (0)\,\mathrm{Col}_3(A) = \mathrm{Col}_1(A) = \langle 3,3,2,1\rangle \\
A\vec{e}_2 &= (0)\,\mathrm{Col}_1(A) + (1)\,\mathrm{Col}_2(A) + (0)\,\mathrm{Col}_3(A) = \mathrm{Col}_2(A) = \langle 0,1,0,0\rangle \\
A\vec{e}_3 &= (0)\,\mathrm{Col}_1(A) + (0)\,\mathrm{Col}_2(A) + (1)\,\mathrm{Col}_3(A) = \mathrm{Col}_3(A) = \langle 1,-1,-1,-2\rangle .
\end{aligned}$$

5) Let A be the matrix given in problem 3 and let I_3 be the 3×3 identity matrix and let I_4 be the 4×4 identity matrix. Verify by computation that $AI_3 = A$ and $I_4A = A$.

Solution:

$$\begin{aligned}
\mathrm{Col}_1(AI_3) &= A\,\mathrm{Col}_1(I_3) = A\vec{e}_1 = \mathrm{Col}_1(A) = \langle 3,3,2,1\rangle \\
\mathrm{Col}_2(AI_3) &= A\,\mathrm{Col}_2(I_3) = A\vec{e}_2 = \mathrm{Col}_2(A) = \langle 0,1,0,0\rangle \\
\mathrm{Col}_3(AI_3) &= A\,\mathrm{Col}_3(I_3) = A\vec{e}_3 = \mathrm{Col}_3(A) = \langle 1,-1,-1,-2\rangle .
\end{aligned}$$

Since all three columns of AI_3 are equal to the corresponding columns of A, then $AI_3 = A$.

$$\begin{aligned}
\mathrm{Row}_1(I_4A) &= A^T\,\mathrm{Row}_1(I_4) = A^T\vec{e}_1 = \mathrm{Col}_1\left(A^T\right) = \mathrm{Row}_1(A) \\
\mathrm{Row}_2(I_4A) &= A^T\,\mathrm{Row}_2(I_4) = A^T\vec{e}_2 = \mathrm{Col}_2\left(A^T\right) = \mathrm{Row}_2(A) \\
\mathrm{Row}_3(I_4A) &= A^T\,\mathrm{Row}_3(I_4) = A^T\vec{e}_3 = \mathrm{Col}_3\left(A^T\right) = \mathrm{Row}_3(A) \\
\mathrm{Row}_4(I_4A) &= A^T\,\mathrm{Row}_4(I_4) = A^T\vec{e}_4 = \mathrm{Col}_4\left(A^T\right) = \mathrm{Row}_4(A) .
\end{aligned}$$

Since all three rows of I_4A are equal to the corresponding rows of A, then $I_4A = A$.

7) Let $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$ be the set of standard unit vectors in R^n. Then

$$\vec{e}_i \cdot \vec{e}_i = \_\_\_\_ \text{ for all } i = 1, 2, \dots, n$$

and

$$\vec{e}_i \cdot \vec{e}_j = \_\_\_\_ \text{ for all } i \text{ and } j \text{ with } i \neq j.$$

Answer:

$$\vec{e}_i \cdot \vec{e}_i = 1 \text{ for all } i = 1, 2, \dots, n$$

and

$$\vec{e}_i \cdot \vec{e}_j = 0 \text{ for all } i \text{ and } j \text{ with } i \neq j.$$

9) Suppose that A is a 3×3 matrix and suppose that B is a 3×3 matrix such that

$$\begin{aligned} A\,\mathrm{Col}_1(B) &= \vec{e}_1 \\ A\,\mathrm{Col}_2(B) &= \vec{e}_2 \\ A\,\mathrm{Col}_3(B) &= \vec{e}_3 \end{aligned}$$

where $\vec{e}_1$, $\vec{e}_2$, and $\vec{e}_3$ are the standard unit vectors in R^3.

Then $AB =$ ________.

Answer:

$$\begin{aligned} \mathrm{Col}_1(AB) &= A\,\mathrm{Col}_1(B) = \vec{e}_1 \\ \mathrm{Col}_2(AB) &= A\,\mathrm{Col}_2(B) = \vec{e}_2 \\ \mathrm{Col}_3(AB) &= A\,\mathrm{Col}_3(B) = \vec{e}_3 \end{aligned}$$

and thus $AB = I_3$.

Exercise 3.7.1
Solution for number 1:
For the matrices

$$A = \begin{bmatrix} 3 & -1 \\ -1 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 3 \\ -1 & -2 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & -3 \\ 2 & 3 \end{bmatrix},$$

we have

$$\begin{aligned} (AB)\,C &= \left(\begin{bmatrix} 3 & -1 \\ -1 & -3 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ -1 & -2 \end{bmatrix} \right) \begin{bmatrix} 1 & -3 \\ 2 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 7 & 11 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & -3 \\ 2 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 29 & 12 \\ 7 & 6 \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} A\,(BC) &= \begin{bmatrix} 3 & -1 \\ -1 & -3 \end{bmatrix} \left(\begin{bmatrix} 2 & 3 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 1 & -3 \\ 2 & 3 \end{bmatrix} \right) \\ &= \begin{bmatrix} 3 & -1 \\ -1 & -3 \end{bmatrix} \begin{bmatrix} 8 & 3 \\ -5 & -3 \end{bmatrix} \\ &= \begin{bmatrix} 29 & 12 \\ 7 & 6 \end{bmatrix}. \end{aligned}$$

This verifies that

$$(AB)\,C = A\,(BC) = \begin{bmatrix} 29 & 12 \\ 7 & 6 \end{bmatrix}.$$

Exercise 3.8.1 **Here are the solutions to numbers 3, 4, and 5.**

3) For

$$A = \begin{bmatrix} 0 & 2 & 3 \\ 3 & -1 & 2 \\ 2 & 2 & 3 \\ 3 & -2 & 3 \end{bmatrix} \quad \text{and} \quad \vec{y} = \langle -1, 1, -3, 4 \rangle,$$

we perform row reduction on the augmented matrix of the equation $A\vec{x} = \vec{y}$ to obtain

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

We see that $A\vec{x} = \vec{y}$ is consistent because the rightmost column of the augmented matrix is not a pivot column. In addition, every column of A is a pivot column so $A\vec{x} = \vec{y}$ has a unique solution. The unique solution is $\vec{x} = \langle -1, -2, 1 \rangle$.

4) For

$$A = \begin{bmatrix} 0 & 2 & 3 \\ 3 & -1 & 2 \\ 2 & 2 & 3 \\ 3 & -2 & 3 \end{bmatrix} \quad \text{and} \quad \vec{y} = \langle 1, -1, -3, 4 \rangle,$$

we perform row reduction on the augmented matrix of the equation $A\vec{x} = \vec{y}$ to obtain

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

We see that $A\vec{x} = \vec{y}$ is inconsistent because the rightmost column of the augmented matrix is a pivot column. The equation $A\vec{x} = \vec{y}$ has no solutions.

5) For

$$A = \begin{bmatrix} -4 & 2 & -2 & -4 \\ 4 & 0 & 0 & 2 \\ 3 & -4 & 2 & -3 \end{bmatrix} \quad \text{and} \quad \vec{y} = \langle -6, 4, 3 \rangle$$

we perform row reduction on the augmented matrix of the equation $A\vec{x} = \vec{y}$ to obtain

$$\left[\begin{array}{cccc|c} -4 & 2 & -2 & -4 & -6 \\ 4 & 0 & 0 & 2 & 4 \\ 3 & -4 & 2 & -3 & 3 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 0 & 0 & \frac{1}{2} & 1 \\ 0 & 1 & 0 & \frac{13}{4} & 1 \\ 0 & 0 & 1 & \frac{17}{4} & 2 \end{array}\right].$$

We see that $A\vec{x} = \vec{y}$ is consistent because the rightmost column of the augmented matrix is not a pivot column. Writing it out in system form, we see that $A\vec{x} = \vec{y}$ is equivalent to the system

$$\begin{array}{ccccccc} x_1 & & & + & \frac{1}{2}x_4 & = & 1 \\ & x_2 & & + & \frac{13}{4}x_4 & = & 1 \\ & & x_3 & + & \frac{17}{4}x_4 & = & 2 \end{array}.$$

We see that x_4 is a free variable (which we can call t) and that the above system has the solution set

$$\begin{aligned} x_1 &= 1 - \frac{1}{2}t \\ x_2 &= 1 - \frac{13}{4}t \\ x_3 &= 2 - \frac{17}{4}t \\ x_4 &= t. \end{aligned}$$

Thus $A\vec{x} = \vec{y}$ has infinitely many solutions and these solutions all have the form

$$\vec{x} = \left\langle 1 - \frac{1}{2}t, 1 - \frac{13}{4}t, 2 - \frac{17}{4}t, t \right\rangle$$

where t can be any real number. We can also write the solution set as

$$\vec{x} = \langle 1, 1, 2, 0 \rangle + t \left\langle \frac{1}{2}, -\frac{13}{4}, -\frac{17}{4}, 1 \right\rangle.$$

As a remark, since t is allowed to be any real number, then we can write $t = 4s$ where s is allowed to be any real number. Then we can write the solution set as

$$\begin{aligned} \vec{x} &= \langle 1, 1, 2, 0 \rangle + (4s) \left\langle \frac{1}{2}, -\frac{13}{4}, -\frac{17}{4}, 1 \right\rangle \\ &= \langle 1, 1, 2, 0 \rangle + s \langle 2, -13, -17, 4 \rangle \end{aligned}$$

where s can be any real number. Many people would rather use as few fractions as possible and the above way of writing the solution set makes it look "nicer" since fractions are not involved.

Exercise 3.8.2

We will answer numbers 2, 4, and 6.

2) Since

$$A = \begin{bmatrix} -3 & -1 \\ 9 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & \frac{1}{3} \\ 0 & 0 \end{bmatrix},$$

we see that A does not have a pivot in every row. This means that there exist some vectors $\vec{y}$ in R^2 such that $A\vec{x} = \vec{y}$ is consistent and there exist other vectors $\vec{y}$ in R^2 for which $A\vec{x} = \vec{y}$ is inconsistent. Not every column of A is a pivot column and this tell us that if $\vec{y}$ is a vector such that $A\vec{x} = \vec{y}$ is consistent, then $A\vec{x} = \vec{y}$ has infinitely many solutions.

4) Since

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

we see that A does not have a pivot in every row. This means that there exist some vectors $\vec{y}$ in R^3 such that $A\vec{x} = \vec{y}$ is consistent and there exist other vectors $\vec{y}$ in R^3 for which $A\vec{x} = \vec{y}$ is inconsistent. Every column of A is a pivot column and this tell us that if $\vec{y}$ is a vector such that $A\vec{x} = \vec{y}$ is consistent, then $A\vec{x} = \vec{y}$ has a unique solution.

6) Since

$$A = \begin{bmatrix} 4 & 1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & \frac{1}{4} & \frac{1}{2} \end{bmatrix},$$

we see that A has a pivot in every row. This tells us that $A\vec{x} = \vec{y}$ is consistent for any choice of vector $\vec{y}$ in R^1. However, not every column of A is a pivot column and this tell us that if $\vec{y}$ is a vector (really a scalar, since R^1 is the set of real numbers) such that $A\vec{x} = \vec{y}$ is consistent, then $A\vec{x} = \vec{y}$ has infinitely many solutions.

Exercise 3.8.3

We will answer numbers 1 and 3.

1) For the matrix

$$A = \begin{bmatrix} 3 & 0 \\ 1 & 3 \end{bmatrix},$$

we have

$$\text{rref}\,(A) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2.$$

This tells us that the equation $A\vec{x} = \vec{y}$ has a unique solution for any choice of vector $\vec{y}$ in R^2.

3) For the matrix

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

we have

$$\text{rref}\,(A) = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -4 \\ 0 & 0 & 0 \end{bmatrix} \neq I_3.$$

This tells us that there are some vectors $\vec{y}$ in R^3 such that $A\vec{x} = \vec{y}$ is inconsistent and that there are also some vectors $\vec{y}$ in R^3 such that $A\vec{x} = \vec{y}$ is consistent. Furthermore, if $\vec{y}$ is a vector such that $A\vec{x} = \vec{y}$ is consistent, then $A\vec{x} = \vec{y}$ has infinitely many solutions.

Exercise 3.8.4:

To solve the equation $AX = I_2$ where

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

we form the augmented matrix

$$\hat{A} = \left[\begin{array}{cc|cc} -1 & -2 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right]$$

and perform row reduction to obtain

$$\text{rref}\left(\hat{A}\right) = \left[\begin{array}{cc|cc} 1 & 0 & -1 & -2 \\ 0 & 1 & 0 & 1 \end{array}\right].$$

Since $\text{rref}\,(A) = I_2$, then $\text{rref}\left(\hat{A}\right) = \left[\begin{array}{c|c} I_2 & X \end{array}\right]$ where X is the unique solution of $AX = I_2$. Thus the unique solution of $AX = I_2$ is

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

It is a coincidence that the solution of $AX = I_2$ happens to be $X = A$.

Exercise 3.9.3:

This is shown in the solution of Exercise 3.8.4.

Exercise 3.9.4:
We will do number 3.
For the matrix

$$A = \begin{bmatrix} \frac{1}{2} & -1 & -6 \\ 0 & 0 & 1 \\ 0 & \frac{1}{2} & 2 \end{bmatrix},$$

to find A^{-1} we form the augmented matrix

$$\hat{A} = \begin{bmatrix} A & I_n \end{bmatrix} = \left[\begin{array}{ccc|ccc} \frac{1}{2} & -1 & -6 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & \frac{1}{2} & 2 & 0 & 0 & 1 \end{array}\right]$$

and then do row reduction on $\hat{A}$ to obtain

$$\operatorname{rref}\left(\hat{A}\right) = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & 4 & 4 \\ 0 & 1 & 0 & 0 & -4 & 2 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{array}\right].$$

Thus

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

Let us check that we got the right answer by computing AA^{-1} (using the A^{-1} we found). We get

$$AA^{-1} = \begin{bmatrix} \frac{1}{2} & -1 & -6 \\ 0 & 0 & 1 \\ 0 & \frac{1}{2} & 2 \end{bmatrix} \begin{bmatrix} 2 & 4 & 4 \\ 0 & -4 & 2 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3,$$

which shows that we have correctly found A^{-1}.

Exercise 3.9.5
We will do number 3.
For the matrix and vector

$$A = \begin{bmatrix} -2 & -2 & 0 \\ 0 & 0 & -3 \\ -1 & 2 & 1 \end{bmatrix} \quad \text{and} \quad \vec{y} = \langle 4, 3, -8 \rangle,$$

we can solve $A\vec{x} = \vec{y}$ by row–reducing the augmented matrix

$$\hat{A} = \left[\begin{array}{c|c} A & \vec{y} \end{array}\right] = \left[\begin{array}{ccc|c} -2 & -2 & 0 & 4 \\ 0 & 0 & -3 & 3 \\ -1 & 2 & 1 & -8 \end{array}\right]$$

to obtain

$$\operatorname{rref}\left(\hat{A}\right) = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & -1 \end{array}\right].$$

This tells us that the unique solution of $A\vec{x} = \vec{y}$ is $\vec{x} = \langle 1, -3, -1\rangle$.

The other way to do this is to first compute

$$A^{-1} = \begin{bmatrix} -\frac{1}{3} & -\frac{1}{9} & -\frac{1}{3} \\ -\frac{1}{6} & \frac{1}{9} & \frac{1}{3} \\ 0 & -\frac{1}{3} & 0 \end{bmatrix}$$

and then observe that the solution of $A\vec{x} = \vec{y}$ is

$$\begin{aligned} \vec{x} &= A^{-1}\vec{y} \\ &= \left\langle \operatorname{Row}_1\left(A^{-1}\right)\cdot\vec{y}, \operatorname{Row}_2\left(A^{-1}\right)\cdot\vec{y}, \operatorname{Row}_3\left(A^{-1}\right)\cdot\vec{y}\right\rangle \\ &= \langle 1, -3, -1\rangle . \end{aligned}$$

Exercise 3.9.7

Suppose that A and B, are $n\times n$ matrices and suppose that A is invertible and B is not invertible. Explain why AB cannot be invertible.

Explanation: Suppose that A is invertible and AB is also invertible. Then since A^{-1} is also invertible and the product of two invertible matrices is invertible, it must be the case that $A^{-1}(AB)$ is invertible. But by the associative law of matrix multiplication,

$$A^{-1}(AB) = \left(AA^{-1}\right)B = I_nB = B,$$

and thus B must be invertible.

We have shown that if A is invertible, then the only possible way to have AB be invertible is to have B be invertible.

Another way to state the conclusion we have arrived it is that if A is invertible and B is not invertible, then AB cannot be invertible.

Exercise 3.9.9

We will do number 1: For the matrix

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

we find by computation that

$$A^{-1} = \begin{bmatrix} \frac{1}{9} & \frac{2}{9} \\ -\frac{5}{9} & -\frac{1}{9} \end{bmatrix}$$

and that

$$A^T = \begin{bmatrix} -1 & 5 \\ -2 & 1 \end{bmatrix}$$

and that

$$\left(A^{-1}\right)^T = \begin{bmatrix} \frac{1}{9} & -\frac{5}{9} \\ \frac{2}{9} & -\frac{1}{9} \end{bmatrix}.$$

Next we observe that

$$\left(A^T\right)\left(A^{-1}\right)^T = \begin{bmatrix} -1 & 5 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{9} & -\frac{5}{9} \\ \frac{2}{9} & -\frac{1}{9} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

and this shows that $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.

Chapter 3 Additional Exercises

1) Complete the following sentences by filling in one of the words "scalar", "vector", or "matrix".

   (a) The sum of two vectors is a **vector**.
   (b) The sum of two matrices is a __________.
   (c) A scalar multiple of a vector is a **vector**.
   (d) A scalar multiple of a matrix is a __________.
   (e) The product of two matrices is a **matrix**.
   (f) The dot product of two vectors is a __________.
   (g) A linear combination of vectors is a **vector**.
   (h) The product of a matrix and a vector is a __________.
   (i) The transpose of a matrix is a __________.

(j) The inverse of a matrix is **matrix**.

3) Suppose that the matrix B has row vectors

$$\begin{aligned}\text{Row}_1(B) &= \langle -6, 4, 4, -2\rangle \\ \text{Row}_2(B) &= \langle 5, -2, 5, 5\rangle \\ \text{Row}_3(B) &= \langle 3, 6, 2, -5\rangle .\end{aligned}$$

Write down B and write down the column vectors of B.

Answer:

$$B = \begin{bmatrix} -6 & 4 & 4 & -2 \\ 5 & -2 & 5 & 5 \\ 3 & 6 & 2 & -5 \end{bmatrix}$$

$$\begin{aligned}\text{Col}_1(B) &= \langle -6, 5, 3\rangle \\ \text{Col}_2(B) &= \langle 4, -2, 6\rangle \\ \text{Col}_3(B) &= \langle 4, 5, 2\rangle \\ \text{Col}_4(B) &= \langle -2, 5, -5\rangle .\end{aligned}$$

5) The $m \times n$ zero matrix is the $m \times n$ matrix that has all entries of 0. This matrix is denoted by $O_{m\times n}$. Thus, for example,

$$O_{3\times 4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Explain why it makes sense to refer to $O_{m\times n}$ as the **additive identity element** for the set of all $m \times n$ matrices.

Answer: It makes sense to call $O_{m\times n}$ as the additive identity element for the set of all $m \times n$ matrices because if A is any $m \times n$ matrix, then $A + O_{m\times n} = A$ (and also $O_{m\times n} + A = A$).

7) For the matrices

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

compute AB and BA.

Solution:

$$AB = \begin{bmatrix} \text{Row}_1(A) \cdot \text{Col}_1(B) & \text{Row}_1(A) \cdot \text{Col}_2(B) \\ \text{Row}_2(A) \cdot \text{Col}_1(B) & \text{Row}_2(A) \cdot \text{Col}_2(B) \end{bmatrix} = \begin{bmatrix} 10 & 4 \\ -7 & -9 \end{bmatrix}$$

and

$BA =$

$$\begin{bmatrix} \text{Row}_1(B) \cdot \text{Col}_1(A) & \text{Row}_1(B) \cdot \text{Col}_2(A) & \text{Row}_1(B) \cdot \text{Col}_3(A) & \text{Row}_1(B) \cdot \text{Col}_4(A) \\ \text{Row}_2(B) \cdot \text{Col}_1(A) & \text{Row}_2(B) \cdot \text{Col}_2(A) & \text{Row}_2(B) \cdot \text{Col}_3(A) & \text{Row}_2(B) \cdot \text{Col}_4(A) \\ \text{Row}_3(B) \cdot \text{Col}_1(A) & \text{Row}_3(B) \cdot \text{Col}_2(A) & \text{Row}_3(B) \cdot \text{Col}_3(A) & \text{Row}_3(B) \cdot \text{Col}_4(A) \\ \text{Row}_4(B) \cdot \text{Col}_1(A) & \text{Row}_4(B) \cdot \text{Col}_2(A) & \text{Row}_4(B) \cdot \text{Col}_3(A) & \text{Row}_4(B) \cdot \text{Col}_4(A) \end{bmatrix}$$

$$= \begin{bmatrix} -6 & -8 & 2 & 4 \\ -4 & -6 & -2 & 4 \\ -12 & -15 & 9 & 6 \\ -2 & -4 & -6 & 4 \end{bmatrix}.$$

9) Suppose that A is an $n \times n$ matrix and let $O_{n\times n}$ be the $n \times n$ zero matrix. (Refer to problem 5 above.) Explain why $AO_{n\times n} = O_{n\times n}$.

Explanation: Since every column vector of $O_{n\times n}$ is equal to $\vec{0}_n$, then for any i and j we have

$$\text{Row}_i(A) \cdot \text{Col}_j(O_{n\times n}) = \text{Row}_i(A) \cdot \vec{0}_n = 0.$$

Thus every entry of $AO_{n\times n}$ is 0. This means that $AO_{n\times n} = O_{n\times n}$.

11) Another property of real numbers that you are probably familiar with is the “cancellation law” which says that if a, b, and c are real numbers with $a \neq 0$ and $ab = ac$, then $b = c$. A similar property *does not,* in general, hold for matrices. Come up with an example of 2×2 matrices A, B, and C such that $A \neq O_{2\times 2}$ and $AB = AC$ but $B \neq C$.

Example: Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 & 2 \\ 4 & -8 \end{bmatrix}, \text{ and } C = \begin{bmatrix} 1 & 2 \\ 3 & -8 \end{bmatrix}.$$

Then $A \neq O_{2\times 2}$ and

$$AB = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 4 & -8 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$$

and

$$AC = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & -8 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$$

so $AB = AC$. However, $B \neq C$.

13) Suppose that A is an $m \times n$ matrix. Explain why the matrix product AA^T is defined (is possible to carry out). What size is AA^T?

Explanation: If A has size $m \times n$, then A^T has size $n \times m$. The number of columns of A and the number of rows of A^T are the same. (Both are n.) Thus AA^T is defined and as size $m \times m$. Also, the number of columns of A^T and the number of rows of A are the same. (Both are m.) Thus $A^T A$ is defined and as size $n \times n$.

15) For the matrix and vector

$$A = \begin{bmatrix} -2 & -1 & 1 & 1 \\ -2 & 2 & -2 & -1 \\ -2 & 2 & 2 & -2 \end{bmatrix} \quad \text{and} \quad \vec{x} = \langle x_1, x_2, x_3, x_4 \rangle \,,$$

Compute $A\vec{x}$ in two different ways: **a)** by using (3.6) and **b)** by using (3.7).

Solution: First we will use (3.6).

$$\begin{aligned} \text{Row}_1\,(A) &= \langle -2, -1, 1, 1 \rangle \\ \text{Row}_2\,(A) &= \langle -2, 2, -2, -1 \rangle \\ \text{Row}_3\,(A) &= \langle -2, 2, 2, -2 \rangle \end{aligned}$$

and thus

$$\begin{aligned} \text{Row}_1\,(A) \cdot \vec{x} &= \langle -2, -1, 1, 1 \rangle \cdot \langle x_1, x_2, x_3, x_4 \rangle \\ &= -2x_1 - x_2 + x_3 + x_4 \end{aligned}$$

and

$$\begin{aligned} \text{Row}_2\,(A) \cdot \vec{x} &= -2x_1 + 2x_2 - 2x_3 - x_4 \\ \text{Row}_3\,(A) \cdot \vec{x} &= -2x_1 + 2x_2 + 2x_3 - 2x_4 \end{aligned}$$

This gives

$$A\vec{x} = \langle -2x_1 - x_2 + x_3 + x_4, -2x_1 + 2x_2 - 2x_3 - x_4, -2x_1 + 2x_2 + 2x_3 - 2x_4 \rangle \,.$$

Now we will use (3.7).

$$\begin{aligned}\operatorname{Col}_1(A) &= \langle -2,-2,-2\rangle\\ \operatorname{Col}_2(A) &= \langle -1,2,2\rangle\\ \operatorname{Col}_3(A) &= \langle 1,-2,2\rangle\\ \operatorname{Col}_4(A) &= \langle 1,-1,-2\rangle .\end{aligned}$$

Thus

$$\begin{aligned}x_1\operatorname{Col}_1(A) &= \langle -2x_1,-2x_1,-2x_1\rangle\\ x_2\operatorname{Col}_2(A) &= \langle -x_2,2x_2,2x_2\rangle\\ x_3\operatorname{Col}_3(A) &= \langle x_3,-2x_3,2x_3\rangle\\ x_4\operatorname{Col}_4(A) &= \langle x_4,-x_4,-2x_4\rangle\end{aligned}$$

and we see that

$$\begin{aligned}A\vec{x} &= x_1\operatorname{Col}_1(A)+x_2\operatorname{Col}_2(A)+x_3\operatorname{Col}_3(A)+x_4\operatorname{Col}_4(A)\\ &= \langle -2x_1-x_2+x_3+x_4,-2x_1+2x_2-2x_3-x_4,-2x_1+2x_2+2x_3-2x_4\rangle .\end{aligned}$$

17) Suppose that A is an $n\times n$ matrix and suppose that the vector $\vec{x}$ in R^n is a solution of the homogeneous equation $A\vec{x}=\vec{0}_n$. Explain why all of the row vectors of A are orthogonal to $\vec{x}$.

Explanation: Note that

$$A\vec{x} = \langle \operatorname{Row}_1(A)\cdot\vec{x},\operatorname{Row}_2(A)\cdot\vec{x},\ldots,\operatorname{Row}_n(A)\cdot\vec{x}\rangle$$

by (3.6). If $A\vec{x}=\vec{0}_n$, then

$$\langle \operatorname{Row}_1(A)\cdot\vec{x},\operatorname{Row}_2(A)\cdot\vec{x},\ldots,\operatorname{Row}_n(A)\cdot\vec{x}\rangle = \langle 0,0,\ldots,0\rangle$$

which means that $\operatorname{Row}_i(A)\cdot\vec{x}=0$ for all $i=1,2,\ldots,n$. We know that two vectors are orthogonal to each other if and only if their dot product is equal to 0. Therefore $\operatorname{Row}_i(A)$ is orthogonal to $\vec{x}$ for all $i=1,2,\ldots,n$.

19) For the following matrices A and vectors $\vec{y}$, find the solution set of the equation $A\vec{x}=\vec{y}$. Indicate whether the equation is inconsistent, consistent with a unique solution, or consistent with infinitely many solutions.

(a)

$$A = \begin{bmatrix} 0 & -1 & 0 \\ -1 & -1 & 0 \\ 1 & -1 & -1 \end{bmatrix} \quad \text{and} \quad \vec{y} = \langle 1, -3, 2 \rangle$$

Solution: The augmented matrix for $A\vec{x} = \vec{y}$ is

$$\hat{A} = \left[\begin{array}{ccc|c} 0 & -1 & 0 & 1 \\ -1 & -1 & 0 & -3 \\ 1 & -1 & -1 & 2 \end{array}\right]$$

and we see that

$$\text{rref}\left(\hat{A}\right) = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

This shows that $A\vec{x} = \vec{y}$ has the unique solution $\vec{x} = \langle 4, -1, 3 \rangle$.

21) Write down the multiply–augmented matrix for the matrix equation $AX = I_2$ where A is the matrix

$$A = \begin{bmatrix} 1 & 4 \\ 3 & 0 \end{bmatrix}$$

and then perform row reduction on this multiply–augmented matrix to find the solution of the equation $AX = I_2$. (What you are doing here is finding A^{-1}.)

Solution: The multiply–augmented matrix for $AX = I_2$ is

$$\hat{A} = \left[\begin{array}{cc|cc} 1 & 4 & 1 & 0 \\ 3 & 0 & 0 & 1 \end{array}\right]$$

and we see that

$$\text{rref}\left(\hat{A}\right) = \left[\begin{array}{cc|cc} 1 & 0 & 0 & \frac{1}{3} \\ 0 & 1 & \frac{1}{4} & -\frac{1}{12} \end{array}\right].$$

This tells us that the unique solution of $AX = I_2$ is

$$X = \begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{4} & -\frac{1}{12} \end{bmatrix}.$$

Since X satisfies $AX = I_2$, then $X = A^{-1}$.

23) Show that the matrix

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

is not invertible by studying the equation $AX = I_3$.

Solution: The multiply–augmented matrix for $AX = I_3$ is

$$\hat{A} = \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{array}\right]$$

and we see that

$$\text{rref}\left(\hat{A}\right) = \left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 0 & -\frac{8}{3} & \frac{5}{3} \\ 0 & 1 & 2 & 0 & \frac{7}{3} & -\frac{4}{3} \\ 0 & 0 & 0 & 1 & -2 & 1 \end{array}\right].$$

Since that fourth column of $\hat{A}$ is $\vec{e}_1$ and this is a pivot column of $\hat{A}$ (as seen after performing the row reduction), then $A\vec{x} = \vec{e}_1$ is inconsistent. This means that $AX = I_3$ is inconsistent. Therefore A is not invertible.

A.4 Chapter 4 Exercises:

Exercise 4.1.1 For each set of one or two vectors, determine whether the set is linearly dependent or linearly independent.

1. $\vec{x} = \langle 1, 0, 1, 2 \rangle$
   Answer: It's nonzero, the set $\{\vec{x}\}$ is linearly independent.

2. $\vec{x}_1 = \langle 1, -1 \rangle$, $\vec{x}_2 = \langle -4, 4 \rangle$
   Answer: The set $\{\vec{x}_1, \vec{x}_2\}$ is linearly dependent, $\vec{x}_2 = -4\vec{x}_1$.

3. $\vec{x}_1 = \langle 0, 0, 0, 0, 0 \rangle$, $\vec{x}_2 = \langle 1, 2, 3, 4, 5 \rangle$
   Answer: The set $\{\vec{x}_1, \vec{x}_2\}$ is linearly dependent, $\vec{x}_1 = 0\vec{x}_2$.

4. $\vec{x}_1 = \langle 3, 6, 18 \rangle$, $\vec{x}_2 = \left\langle \frac{1}{3}, \frac{2}{3}, 2 \right\rangle$
   Answer: The set $\{\vec{x}_1, \vec{x}_2\}$ is linearly dependent, $\vec{x}_1 = 9\vec{x}_2$

5. $\vec{x}_1 = \langle 2, 1, 0\rangle$, $\vec{x}_2 = \langle -1, 2, 0\rangle$
   Answer: The set $\{\vec{x}_1, \vec{x}_2\}$ is linearly independent. The vectors are not scalar multiplies of each other.

Exercise 4.1.2 For each matrix A, determine if the columns are linearly dependent or linearly independent. If dependent, find a linear dependence relation.

1. $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$
   Answer: Note that $\operatorname{rref}(A) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$. From this, we see that the columns are linearly dependent, and
   $$\operatorname{Col}_3(A) = -\operatorname{Col}_1(A) + 2\operatorname{Col}_2(A).$$
   We can rearrange to get a linear dependence relation
   $$\operatorname{Col}_1(A) - 2\operatorname{Col}_2(A) + \operatorname{Col}_3(A) = \vec{0}_3.$$

2. $A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 5 & 1 \\ 1 & 3 & 1 \end{bmatrix}$
   Answer: $\operatorname{rref}(A) = I_3$, so the columns are linearly independent.

Exercise 4.1.3 Without performing any computations explain why each of the following sets of vectors is linearly dependent.

1. $\{\langle 1, 0, 0\rangle, \langle 0, 1, 0\rangle, \langle 0, 0, 0\rangle\}$
   Answer: It includes the zero vector.

2. $\{\langle 1, 2\rangle, \langle 2, 1\rangle, \langle 3, 0\rangle\}$
   Answer: There are three vectors in R^2 (more vectors than components in each).

3. $\{\langle 1, 0, 0\rangle, \langle 0, 1, 0\rangle, \langle 0, 0, 1\rangle, \langle 1, -1, 1\rangle\}$
   Answer: Similarly, there are four vectors in R^3.

4. $\{\langle 1,2,1\rangle, \langle -1,-2,-1\rangle, \langle 1,0,0\rangle\}$
   Answer: The second vector is a multiple of the first.

Exercise 4.1.4 For each set of vectors, determine if the set is linearly dependent or linearly independent. If dependent, find a linear dependence relation.

1. $\{\langle 1,1,2\rangle, \langle 2,-1,0\rangle, \langle 1,-3,1\rangle\}$
   Answer: They are independent. If we create a matrix with these as columns, it row reduces to I_3

2. $\{\langle 1,2,0,-3\rangle, \langle -2,-4,0,6\rangle, \langle 0,2,3,1\rangle, \langle 1,6,6,-1\rangle\}$
   Answer: They are linearly dependent. This can be found by creating a matrix having these vectors as its column vectors. If we call the vectors $\vec{x}_1, \vec{x}_2, \vec{x}_3$ and $\vec{x}_4$ in the order they're given in, then a linear dependence relation is any equation of the form

$$(2s-t)\vec{x}_1 + s\vec{x}_2 - 2t\vec{x}_3 + t\vec{x}_4 = \vec{0}_4,$$

   with at least one of s or t being nonzero.

3. $\{\langle 0,4,-2,5\rangle, \langle 3,7,-5,-4\rangle, \langle 1,5,-3,2\rangle\}$
   Answer: They are linearly dependent. Calling the vectors $\vec{x}_1$, $\vec{x}_2$ and $\vec{x}_3$, in the order given, a linear dependence relation is any equation of the form

$$2t\vec{x}_1 + t\vec{x}_2 - 3t\vec{x}_3 = \vec{0}_4$$

   for $t \neq 0$.

4. $\{\langle 3,1,0,-1\rangle, \langle 2,0,-2,8\rangle, \langle 3,1,5,4\rangle\}$
   Answer: These are linearly independent. Again, set up a matrix with these as its columns, and do row reduction. All three columns will be pivot columns.

Exercise 4.2.1 Determine whether each subset is a subspace of R^n for the indicated value of n.
Details shown for 2., 4., and 5.

1. $S = \{\langle 0,a\rangle \in R^2 \,|\, a \in R\}$
   Answer: This is a subspace. All three criteria can be established.

2. $T = \{\langle 1, a\rangle \in R^2 \,|\, a \in R\}$
   Answer: This is not a subspace. Note $\vec{x} = \langle 1, 0\rangle$ is in T, but $2\vec{x} = \langle 2, 0\rangle$ is not in T. So T is not closed under scalar multiplication. (It's not closed under vector addition either.)

3. $Q = \{\langle 0, 0, 0\rangle\}$ in R^3
   Answer: This is a subspace. All three criteria can be established.

4. $P = \{\langle k, k\rangle \in R^2 \,|\, k = 1, 2, \dots\}$
   Answer: This is not a subspace. For example, $\vec{x} = \langle 1, 1\rangle$ is in P, but $\frac{1}{2}\vec{x} = \left\langle \frac{1}{2}, \frac{1}{2}\right\rangle$ is not in P. P is not closed under scalar multiplication. Incidentally, P is closed under vector addition.

5. $L = \{\langle k, k\rangle \in R^2 \,|\, k \in R\}$
   Answer: This is a subspace. Here is a detailed justification for this conclusion. First, L is nonempty, for example it contains the vector $\langle 1, 1\rangle$. We should note that the property of vectors in L is that their first and second entries are the same real number. We need to show that this property is preserved when we add such vectors or scale them. Suppose $\vec{x}_1 = \langle k_1, k_1\rangle$ and $\vec{x}_2 = \langle k_2, k_2\rangle$ are any vectors in L, and let c be any scalar. Note that

$$\vec{x}_1 + \vec{x}_2 = \langle k_1 + k_2, k_1 + k_2\rangle,$$

   and

$$c\vec{x}_1 = \langle ck_1, ck_1\rangle.$$

   Note that $\vec{x}_1 + \vec{x}_2$ and $c\vec{x}_1$ satisfy the property necessary to be elements of L. Hence L is a subspace of R^2. We can also argue that $L = \text{Span}\{\langle 1, 1\rangle\}$, and as a span, it is necessarily a subspace.

Exercise 4.2.2 Find a spanning set for each subspace of R^n.
Details shown for 3.

- $Q = \{\langle 0, a\rangle \in R^2 \,|\, a \in R\}$
  Answer: $Q = \text{Span}\{\langle 0, 1\rangle\}$

- $P = \{\langle a, a, b\rangle \in R^3 \,|\, a, b \in R\}$
  Answer: $P = \text{Span}\{\langle 1, 1, 0\rangle, \langle 0, 0, 1\rangle\}$

- $T = \{\langle a, b, c, a+b+c\rangle \in R^4 \mid a, b, c \in R\}$
  Answer: $T = \text{Span}\{\langle 1,0,0,1\rangle, \langle 0,1,0,1\rangle, \langle 0,0,1,1\rangle\}$. Note that an element of T can be decomposed.

$$\begin{aligned}\langle a, b, c, a+b+c\rangle &= \langle a,0,0,a\rangle + \langle 0,b,0,b\rangle + \langle 0,0,c,c\rangle \\ &= a\langle 1,0,0,1\rangle + b\langle 0,1,0,1\rangle + c\langle 0,0,1,1\rangle\end{aligned}$$

Exercise 4.2.3 Consider the set of vectors

$$T = \{\langle 1,0,1,1\rangle, \langle -2,3,0,8\rangle, \langle 4,4,5,2\rangle\}.$$

1. Find a matrix A having row space $\mathcal{RS}(A) = \text{Span}(T)$.
   Answer: Answers can vary, but the obvious choice is the matrix whose row vectors are the elements of A.

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

2. Find a matrix A having columns space $\mathcal{CS}(A) = \text{Span}(T)$.
   Answer: Answers can vary, but the obvious choice is the matrix whose column vectors are the elements of A.

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

Exercise 4.2.4 Find a spanning set for the null space $\mathcal{N}(A)$ where

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

Answer: The solutions to $A\vec{x} = \vec{0}_4$ are given in vector parametric form $\vec{x} = t\left\langle -\frac{2}{3}, -\frac{1}{3}, 1\right\rangle$. So $\mathcal{N}(A) = \text{Span}\left\{\left\langle -\frac{2}{3}, -\frac{1}{3}, 1\right\rangle\right\}$.

Exercise 4.2.5 For each matrix A, find a spanning set for each of the four fundamental subspaces of A.

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

   Answer:

$$\begin{aligned} \mathcal{RS}(A) &= \text{Span}\{\langle 1,2,0,1\rangle, \langle 2,1,3,0\rangle, \langle 3,-1,7,3\rangle, \langle 2,0,4,0\rangle\}, \\ \mathcal{CS}(A) &= \text{Span}\{\langle 1,2,3,2\rangle, \langle 2,1,-1,0\rangle, \langle 0,3,7,4\rangle, \langle 2,0,3,0\rangle\}, \\ \mathcal{N}(A) &= \text{Span}\left\{\langle -2,1,1,0\rangle\right\}, \quad \text{and} \\ \mathcal{N}(A^T) &= \text{Span}\left\{\langle 6/19,-16/19,-4/19,1\rangle\right\}. \end{aligned}$$

2. $A = \begin{bmatrix} 1 & 3 & 1 & 0 & 11 \\ -1 & 1 & 4 & 1 & 0 \\ -2 & 0 & 3 & -1 & -3 \end{bmatrix}$

   Answer:

$$\begin{aligned} \mathcal{RS}(A) &= \text{Span}\{\langle 1,3,1,0,11\rangle, \langle -1,1,4,1,0\rangle, \langle -2,0,3,-1,-3\rangle\}, \\ \mathcal{CS}(A) &= \text{Span}\{\langle 1,-1,-2\rangle, \langle 3,1,0\rangle, \langle 1,4,3\rangle, \langle 0,1,-1\rangle, \langle 11,0,-3\rangle\}, \\ \mathcal{N}(A) &= \text{Span}\left\{\langle -2,1,-1,1,0\rangle, \langle 0,-4,1,0,1\rangle\right\}, \quad \text{and} \\ \mathcal{N}(A^T) &= \text{Span}\left\{\langle 0,0,0\rangle\right\}. \end{aligned}$$

3. $A = \begin{bmatrix} 4 & 8 & -3 & 1 \\ -2 & -4 & 5 & -11 \\ 3 & 6 & 1 & -9 \end{bmatrix}$

   Answer:

$$\begin{aligned} \mathcal{RS}(A) &= \text{Span}\{\langle 4,8,-3,1\rangle, \langle -2,-4,5,-11\rangle, \langle 3,6,1,-9\rangle\}, \\ \mathcal{CS}(A) &= \text{Span}\{\langle 4,-2,3\rangle, \langle 8,-4,6\rangle, \langle -3,5,1\rangle, \langle 1,-11,-9\rangle\}, \\ \mathcal{N}(A) &= \text{Span}\left\{\langle -2,1,0,0\rangle, \langle 2,0,3,1\rangle\right\}, \quad \text{and} \\ \mathcal{N}(A^T) &= \text{Span}\left\{\langle -17/14,-13/14,1\rangle\right\}. \end{aligned}$$

Exercise 4.3.1 Suppose S is a subspace of R^n for some $n \geq 2$, and let $\mathcal{B} = \{\vec{u}_2, \ldots, \vec{u}_k\}$ be a basis for S. Explain why the number of vectors, k, in the basis $\mathcal{B}$ must be less than or equal to n.

Answer: We know that if a set contains more vectors than there are components in each vector, the set is linearly dependent (see Theorem 4.1.2). The basis $\mathcal{B}$ is a linearly independent set of vectors in R^n (each of its elements has n components), so it contains at most n vectors.

Exercise 4.3.2 Find a basis for each subspace of R^n. You may wish to start with the spanning sets you found for these same subspaces in Exercise 4.2.2, but you should demonstrate that the set you claim is a basis is both a spanning set and is linearly independent.
Details shown for 3.

- $Q = \{\langle 0, a\rangle \in R^2 \mid a \in R\}$
  Answer: Calling the basis $\mathcal{B}$, $\mathcal{B} = \{\langle 0, 1\rangle\}$.

- $P = \{\langle a, a, b\rangle \in R^3 \mid a, b \in R\}$
  Answer: Calling the basis $\mathcal{B}$, $\mathcal{B} = \{\langle 1, 1, 0\rangle, \langle 0, 0, 1\rangle\}$.

- $T = \{\langle a, b, c, a + b + c\rangle \in R^4 \mid a, b, c \in R\}$
  Answer: Calling the basis $\mathcal{B}$, $\mathcal{B} = \{\langle 1, 0, 0, 1\rangle, \langle 0, 1, 0, 1\rangle, \langle 0, 0, 1, 1\rangle\}$. It is a spanning set since we can write any element $\langle a, b, c, a+b+c\rangle$ of T as a linear combination

  $$\langle a, b, c, a + b + c\rangle = a\langle 1, 0, 0, 1\rangle + b\langle 0, 1, 0, 1\rangle + c\langle 0, 0, 1, 1\rangle.$$

  To show that they are linearly independent, consider the matrix A having these as its column vectors.

  $$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \xrightarrow{rref} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

  Since all columns are pivot columns, the columns of A, the vectors in our set $\mathcal{B}$, are linearly independent.

Exercise 4.3.3 Find a basis for $\mathcal{N}(A)$ and for $\mathcal{N}(A^T)$ for each matrix A.

1. $A = \begin{bmatrix} 1 & 3 & 0 & 1 \\ 2 & 6 & 0 & 2 \\ 3 & 9 & 1 & 2 \end{bmatrix}$
   Answer: A basis for $\mathcal{N}(A)$ is $\{\langle -3, 1, 0, 0\rangle, \langle -1, 0, 1, 1\rangle\}$ and a basis for $\mathcal{N}(A^T)$ is $\{\langle -2, 1, 0\rangle\}$.

2. $A = \begin{bmatrix} 2 & 3 & 1 \\ -3 & 5 & 8 \\ 4 & 2 & -2 \end{bmatrix}$
   Answer: A basis for $\mathcal{N}(A)$ is $\{\langle 1, -1, 1\rangle\}$ and a basis for $\mathcal{N}(A^T)$ is $\left\{\left\langle -\frac{26}{19}, \frac{8}{19}, 1\right\rangle\right\}$.

Exercise 4.3.4 Consider the basis $\mathcal{B} = \{\langle 1, 1\rangle, \langle 1, -1\rangle\}$ of R^2.

1. Find $[\vec{x}]_{\mathcal{B}}$ if $\vec{x} = \langle 1, 1\rangle$
   Answer: $[\vec{x}]_{\mathcal{B}} = \langle 1, 0\rangle$

2. Find $[\vec{x}]_{\mathcal{B}}$ if $\vec{x} = \langle 1, -1\rangle$
   Answer: $[\vec{x}]_{\mathcal{B}} = \langle 0, 1\rangle$

3. Find $[\vec{x}]_{\mathcal{B}}$ if $\vec{x} = \langle 0, 0\rangle$
   Answer: $[\vec{x}]_{\mathcal{B}} = \langle 0, 0\rangle$

4. Find $[\vec{x}]_{\mathcal{B}}$ if $\vec{x} = \langle 2, 3\rangle$
   Answer: $[\vec{x}]_{\mathcal{B}} = \left\langle \frac{5}{2}, -\frac{1}{2}\right\rangle$

5. Find $\vec{x}$ if $[\vec{x}]_{\mathcal{B}} = \langle -1, 4\rangle$
   Answer: $\vec{x} = -1\langle 1, 1\rangle + 4\langle 1, -1\rangle = \langle 3, -5\rangle$.

Exercise 4.3.5 For the subspace P in Example 4.3.9, construct the matrix B whose columns are the basis elements in the order given. For each coordinate vector $[\vec{x}]_{\mathcal{B}}$ in R^2, find the element $\vec{x}$ in P by using the matrix-vector product $\vec{x} = B[\vec{x}]_{\mathcal{B}}$.

Details: The matrix $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$.

1. $[\vec{x}]_{\mathcal{B}} = \langle 1, 1\rangle$
   Answer: $B[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$. So $\vec{x} = \langle 1, 0, 1, 0\rangle$.

2. $[\vec{x}]_{\mathcal{B}} = \langle -3, 5\rangle$

   Answer: $B[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -3 \\ 5 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 5 \\ 0 \end{bmatrix}$. So $\vec{x} = \langle -3, 0, 5, 0\rangle$.

3. $[\vec{x}]_{\mathcal{B}} = \langle x_1, x_2\rangle$

   Answer: $B[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ 0 \\ x_2 \\ 0 \end{bmatrix}$. So $\vec{x} = \langle x_1, 0, x_2, 0\rangle$.

Exercise 4.3.6 Consider the vectors $\vec{u}_1$, $\vec{u}_2$, and $\vec{u}_3$ in R^5 given by

$$\vec{u}_1 = \langle 1,0,0,0,0\rangle, \quad \vec{u}_2 = \langle 1,1,0,0,0\rangle, \quad \text{and} \quad \vec{u}_3 = \langle 1,1,1,0,0\rangle.$$

Let $\mathcal{B} = \{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ be the ordered basis for the subspace $S = \text{Span}(\mathcal{B})$ of R^5. Verify that the coordinate vectors, $[\vec{u}_i]_{\mathcal{B}}$, of the basis elements are the standard unit vectors in R^3. That is, show that

$$[\vec{u}_1]_{\mathcal{B}} = \langle 1,0,0\rangle, \quad [\vec{u}_2]_{\mathcal{B}} = \langle 0,1,0\rangle, \quad \text{and} \quad [\vec{u}_3]_{\mathcal{B}} = \langle 0,0,1\rangle.$$

Answer: This can be done using a matrix with the basis elements as column vectors, or it can be demonstrated by considering the coefficients when expressing one of the basis elements as a linear combination. Note that if $\vec{u}_1 = c_1\vec{u}_1 + c_2\vec{u}_2 + c_3\vec{u}_3$, then $[\vec{u}_1]_{\mathcal{B}} = \langle c_1, c_2, c_3\rangle$. Since

$$\vec{u}_1 = 1\vec{u}_1 + 0\vec{u}_2 + 0\vec{u}_3,$$

the coefficients for $\vec{u}_1$ as a linear combination of the basis elements are $c_1 = 1$, $c_2 = 0$, and $c_3 = 0$. This makes

$$[\vec{u}_1]_{\mathcal{B}} = \langle 1,0,0\rangle.$$

Similar observations lead to $[\vec{u}_2]_{\mathcal{B}} = \langle 0,1,0\rangle$, and $[\vec{u}_3]_{\mathcal{B}} = \langle 0,0,1\rangle$. To verify using a matrix, we let B have column vectors from the basis elements (in the given order).

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

Note that if we compute $B\vec{e}_1$, for the standard unit vector $\vec{e}_1 = \langle 1, 0, 0\rangle$ in R^3, we get

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

which is $\vec{u}_1$ from our basis. Similar computations show that $B\vec{e}_2 = \vec{u}_2$ and $B\vec{e}_3 = \vec{u}_3$.

Exercise 4.3.7 Suppose $n \geq 2$ and $1 \leq k \leq n$. Let $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_k\}$ be an ordered basis of the subspace $S = \text{Span}(\mathcal{B})$ of R^n. Explain why $[\vec{u}_i]_{\mathcal{B}} = \vec{e}_i$ for each $i = 1, \ldots, k$. That is, explain why the coordinate vectors for the basis elements are the standard unit vectors in R^k.

Hint: don't worry about a bunch of computations, just consider the equation

$$\vec{u}_i = c_1\vec{u}_1 + \cdots + c_i\vec{u}_i + \cdots + c_k\vec{u}_k.$$

Answer: To express $\vec{u}_i$ as a linear combination of the basis elements, we need the i^{th} coefficient to be 1 and all the other coefficients to be zero. For example,

$$\vec{u}_1 = 1\vec{u}_1 + 0\vec{u}_2 + \cdots + 0\vec{u}_k,$$

making $[\vec{u}_1]_{\mathcal{B}} = \langle 1, 0, \ldots, 0\rangle = \vec{e}_1$, and so forth. Note that since there are k basis elements, this coordinate vector, $\vec{e}_1$, has k entries—one 1 and $k-1$ zeros.

Exercise 4.3.8 Find the dimension of the null space of the matrix

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

Answer: To identify the solution of $A\vec{x} = \vec{0}_4$, we row reduce $\left[A \mid \vec{0}_4\right]$.

$$\left[A \mid \vec{0}_4\right] \quad \xrightarrow{rref} \quad \left[\begin{array}{ccccc|c} 1 & 0 & -3 & 0 & 1 & 0 \\ 0 & 1 & 4 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

For $\vec{x} = \langle x_1, x_2, x_3, x_4, x_5 \rangle$,

$$\begin{array}{rcl} x_1 & = & 3x_3 - x_5 \\ x_2 & = & -4x_3 + 2x_5 \\ x_4 & = & -2x_5 \end{array},$$

with x_3 and x_5 free variables. So a solution to the homogeneous equation has the form

$$\vec{x} = x_3\langle 3, -4, 1, 0, 0 \rangle + x_5\langle -1, 2, 0, -2, 1 \rangle.$$

The set $\{\langle 3, -4, 1, 0, 0 \rangle, \langle -1, 2, 0, -2, 1 \rangle\}$ is a basis and it contains two vectors. So

$$\dim\big(\mathcal{N}(A)\big) = 2.$$

Exercise 4.3.9 Find the dimension of the null space of the matrix

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

Answer: Consider $B\vec{x} = \vec{0}_5$.

$$\big[B \mid \vec{0}_5\big] \quad \overset{rref}{\longrightarrow} \quad \left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

A solution $\vec{x} = \langle x_1, x_2, x_3, x_4 \rangle$, satisfies

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

with x_4 a free variable. So the vectors in $\mathcal{N}(B)$ have the form

$$\vec{x} = x_4\langle -1, 1, 1, 1 \rangle,$$

and a basis is $\{\langle -1,1,1,1\rangle\}$, containing one vector.

$$\dim\big(\mathcal{N}(B)\big) = 1.$$

Exercise 4.3.10 Consider the matrix H.

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

1. Classify each column as a pivot column or a non-pivot column.
   Answer: The pivot columns are the first, third and fourth, and the non-pivot columns are the second and fifth.

2. Express each non-pivot column as a linear combination of one or more pivot columns.
   Answer: Since H is an rref,

$$\begin{aligned} \mathrm{Col}_2(H) &= -2\,\mathrm{Col}_1(H), \\ \mathrm{Col}_5(H) &= 3\,\mathrm{Col}_1(H) - 4\,\mathrm{Col}_3(H) + 5\,\mathrm{Col}_4(H). \end{aligned}$$

3. Identify a basis for $\mathcal{CS}(H)$.
   Answer: A basis is the set of pivot columns, $\{\langle 1,0,0,0\rangle, \langle 0,1,0,0\rangle, \langle 0,0,1,0\rangle\}$.

Exercise 4.4.1 Find a basis for the column space of each matrix.

1. $A = \begin{bmatrix} 3 & 3 & 3 \\ -1 & -2 & 1 \\ -5 & -6 & -3 \end{bmatrix}$
   Answer: $\{\langle 3,-1,-5\rangle, \langle 3,-2,-6\rangle\}$

2. $M = \begin{bmatrix} 1 & 3 & 1 & 0 & 2 \\ 2 & 2 & -2 & 4 & 0 \\ 3 & 1 & -5 & 8 & 1 \end{bmatrix}$
   Answer: $\{\langle 1,2,3\rangle, \langle 3,2,1\rangle, \langle 2,0,1\rangle\}$

3. $X = \begin{bmatrix} 1 & 2 & -1 & 4 \\ 0 & 0 & 3 & -3 \\ 2 & 4 & 1 & 5 \\ 1 & 2 & 1 & 2 \\ 3 & 6 & 2 & 7 \end{bmatrix}$
   Answer: $\{\langle 1,0,2,1,3\rangle, \langle -1,3,1,1,2\rangle\}$

Exercise 4.4.2 Find bases for the row space, column space, and null space of each matrix.

1. $A = \begin{bmatrix} 3 & 3 & 3 \\ -1 & -2 & 1 \\ -5 & -6 & -3 \end{bmatrix}$
   Answer: Column $\{\langle 3,-1,-5\rangle, \langle 3,-2,-6\rangle\}$,
   Row $\{\langle 1,0,3\rangle, \langle 0,1,-2\rangle\}$,
   null $\{\langle -3,2,1\rangle\}$

2. $M = \begin{bmatrix} 1 & 3 & 1 & 0 & 2 \\ 2 & 2 & -2 & 4 & 0 \\ 3 & 1 & -5 & 8 & 1 \end{bmatrix}$
   Answer: Column $\{\langle 1,2,3\rangle, \langle 3,2,1\rangle, \langle 2,0,1\rangle\}$,
   Row $\{\langle 1,0,-2,3,0\rangle, \langle 0,1,1,-1,0\rangle, \langle 0,0,0,0,1\rangle\}$,
   null $\{\langle 2,-1,1,0,0\rangle, \langle -1,1,0,1,0\rangle\}$

3. $X = \begin{bmatrix} 1 & 2 & -1 & 4 \\ 0 & 0 & 3 & -3 \\ 2 & 4 & 1 & 5 \\ 1 & 2 & 1 & 2 \\ 3 & 6 & 2 & 7 \end{bmatrix}$
   Answer: Column $\{\langle 1,0,2,1,3\rangle, \langle -1,3,1,1,2\rangle\}$,
   Row $\{\langle 1,2,0,3\rangle, \langle 0,0,1,-1\rangle\}$,
   null $\{\langle -2,1,0,0\rangle, \langle -3,0,1,1\rangle\}$

Exercise 4.5.1 Suppose A is a 10×20 matrix.

1. If $\text{rank}(A) = 7$, what is $\text{nullity}(A)$?
   Answer: $\text{nullity}(A) = 20 - 7 = 13$

2. If $\text{rank}(A) = 7$, what is $\text{nullity}(A^T)$?
   Answer: $\text{nullity}(A^T) = 10 - 7 = 3$

3. If A^T has 8 pivot columns, what is rank(A)?
   Answer: $\text{rank}(A) = \text{rank}(A^T) = 8$

4. If $\dim\big(\mathcal{CS}(A^T)\big) = 9$, how many free variables are there in any solution to $A\vec{x} = \vec{0}_{10}$?
   Answer: $\text{rank}(A) = \text{rank}(A^T) = \dim\big(\mathcal{CS}(A^T)\big) = 9$. So $\text{nullity}(A) = 20 - \text{rank}(A) = 20 - 9 = 11$. There are eleven free variables (corresponding to the 11 non-pivot columns).

5. What is the maximum possible rank of A?
   Answer: The maximum rank if 10. There are at most 10 pivot positions since there are ten rows.

Exercise 4.5.2 Explain why the maximum rank of an $m \times n$ matrix is the smaller of the two numbers, m and n.
Answer: The rank of a matrix is the number of pivot columns which is the same as the number of pivot positions. Each pivot position is in a row and a column, so the rank can't exceed the smallest of these two numbers.

Exercise 4.5.3 Suppose A is an $n \times n$ matrix (so A is square). Explain why if A is full rank, then $\text{nullity}(A) = 0$.
Answer: Since A is square the maximum rank is n. If A is full rank, then $\text{rank}(A) = n$ meaning that by the rank-nullity theorem $\text{nullity}(A) = n - \text{rank}(A) = n - n = 0$.

Exercise 4.6.1: Verify that $V = \left\{\vec{0}\right\}$ with operations as defined above satisfies all of the axioms given in Definition 4.6.1 (and is thus a vector space).

Proof:

1. The only vector in V is $\vec{0}$ and $\vec{0} + \vec{0} = \vec{0} \in V$, so V is closed under addition.

2. The only vector in V is $\vec{0}$ and if c is any scalar then we have $c\vec{0} = \vec{0} \in V$, so V is closed under scalar multiplication.

3. $\vec{0} + \vec{0} = \vec{0} + \vec{0}$, so V satisfies the commutative property of addition.

4. $\left(\vec{0}+\vec{0}\right)+\vec{0}=\vec{0}+\vec{0}=\vec{0}$ and $\vec{0}+\left(\vec{0}+\vec{0}\right)=\vec{0}+\vec{0}=\vec{0}$, so V satisfies the associative property of addition.

5. $\vec{0}$ is clearly the zero vector of V.

6. It is clear that $\vec{0}$ is the additive inverser of $\vec{0}$ because $\vec{0}+\vec{0}=\vec{0}$.

7. If c is any scalar then $c\left(\vec{0}+\vec{0}\right)=c\vec{0}=\vec{0}$ and $c\vec{0}+c\vec{0}=\vec{0}+\vec{0}=\vec{0}$, so the distributive property is satisfied.

8. If c and d are any two scalars, then $(c+d)\,\vec{0}=\vec{0}$ and $c\vec{0}+d\vec{0}=\vec{0}+\vec{0}=\vec{0}$. Thus $(c+d)\,\vec{0}=c\vec{0}+d\vec{0}$.

9. If c and d are any two scalars then $(cd)\,\vec{0}=\vec{0}$ and $c\left(d\vec{0}\right)=c\vec{0}=\vec{0}$ so it is true that $(cd)\,\vec{0}=c\left(d\vec{0}\right)$.

10. $(1)\,\vec{0}=\vec{0}$ by our definition of scalar multiplication. (Any scalar times $\vec{0}$ is $\vec{0}$ by definition.)

Exercise 4.6.3: Prove statements 2 and 4 of Theorem 4.6.1.

Proof of statement 2: We want to prove that if V is a vector space $\vec{x}$ is any vector in V, then there is only one vector that can serve as an additive inverse for $\vec{x}$.

Suppose that $\vec{x}\in V$ and suppose that there are vectors $\vec{y}$ and $\vec{z}$ in V that both serve as additive inverses for $\vec{x}$. This means that $\vec{x}+\vec{y}=\vec{0}$ and $\vec{x}+\vec{z}=\vec{0}$. Hence $\vec{x}+\vec{y}=\vec{x}+\vec{z}$. We can add $\vec{z}$ to both sides of this equation to obtain

$$\vec{z}+(\vec{x}+\vec{y})=\vec{z}+(\vec{x}+\vec{z})$$

and then use the associative property to obtain

$$(\vec{z}+\vec{x})+\vec{y}=(\vec{z}+\vec{x})+\vec{z}.$$

This gives

$$\vec{0}+\vec{y}=\vec{0}+\vec{z}$$

and (because $\vec{0}$ is the additive identity element)

$$\vec{y}=\vec{z}.$$

We have proved that $\vec{x}$ cannot have two different additive inverses.

Proof of statement 4: We want to prove that if V is a vector space and c is any scalar, then $c\vec{0} = \vec{0}$.

First note that one of the axioms tells us that we must have

$$c\left(\vec{0} + \vec{0}\right) = c\vec{0} + c\vec{0}.$$

Since $\vec{0} + \vec{0} = \vec{0}$ (because $\vec{0}$ is the additive identity element), then we have

$$c\vec{0} = c\vec{0} + c\vec{0}.$$

We know that the vector $c\vec{0}$ has an additive inverse, denoted by $-\left(c\vec{0}\right)$. If we add this to both sides of the above equation, we obtain

$$c\vec{0} + \left(-\left(c\vec{0}\right)\right) = \left(c\vec{0} + c\vec{0}\right) + \left(-\left(c\vec{0}\right)\right).$$

We then use the associative property to obtain

$$c\vec{0} + \left(-\left(c\vec{0}\right)\right) = c\vec{0} + \left(c\vec{0} + \left(-\left(c\vec{0}\right)\right)\right).$$

This gives

$$\vec{0} = c\vec{0} + \vec{0}$$

and since $c\vec{0} + \vec{0} = c\vec{0}$, we have $c\vec{0} = \vec{0}$, which is what we wanted to prove.

Exercise 4.7.1 Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ be any element of $M_{2\times 2}$. Show that $-1A$ is the additive inverse of A.

Answer: We have to show that $-1A + A = O$. This is what it means to be the additive inverse. Using the operations defined back in section 3.2

$$\begin{aligned} -1A + A &= \begin{bmatrix} -1a_{11} & -1a_{12} \\ -1a_{21} & -1a_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\ &= \begin{bmatrix} -1a_{11} + a_{11} & -1a_{12} + a_{12} \\ -1a_{21} + a_{21} & -1a_{22} + a_{22} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

Hence $-1A$ is the additive inverse $-A$.

Exercise 4.7.5 Consider the ordered basis $\mathcal{B} = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ of $M_{2\times 2}$ where

$$E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

1. If A is any matrix in $M_{2\times 2}$, then the coordinate vector $[A]_{\mathcal{B}}$ is a vector in R^k. Determine the value of k.
   Answer: $k = 4$ (since there are four basis vectors)

2. Find the coordinate vectors $[A]_{\mathcal{B}}$ and $[B]_{\mathcal{B}}$ for the matrices

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

   Answer: $[A]_{\mathcal{B}} = \langle 2, -4, 3, 1\rangle$ and $[B]_{\mathcal{B}} = \langle -3, 2, 0, 5\rangle$

3. Evaluate $A + B$ and confirm that $[A]_{\mathcal{B}} + [B]_{\mathcal{B}} = [A + B]_{\mathcal{B}}$
   Answer: $A + B = \begin{bmatrix} -1 & -2 \\ 3 & 6 \end{bmatrix}$, so $[A + B]_{\mathcal{B}} = \langle -1, -2, 3, 6\rangle$. Also $[A]_{\mathcal{B}} + [B]_{\mathcal{B}} = \langle 2, -4, 3, 1\rangle + \langle -3, 2, 0, 5\rangle = \langle -1, -2, 3, 6\rangle$. As expected, they are the same.

4. Evaluate $5A$ and confirm that $5[A]_{\mathcal{B}} = [5A]_{\mathcal{B}}$.
   Answer: $5A = \begin{bmatrix} 10 & -20 \\ 15 & 5 \end{bmatrix}$, so that $[5A]_{\mathcal{B}} = \langle 10, -20, 15, 5\rangle$. Also $5[A]_{\mathcal{B}} = 5\langle 2, -4, 3, 1\rangle = \langle 10, -20, 15, 5\rangle$. Again, they match.

5. Find the coordinate vectors for the elements of $\mathcal{B}$. That is, find each of $[E_{11}]_{\mathcal{B}}$, $[E_{12}]_{\mathcal{B}}$, $[E_{21}]_{\mathcal{B}}$, and $[E_{22}]_{\mathcal{B}}$.
   Answer: $[E_{11}]_{\mathcal{B}} = \langle 1, 0, 0, 0\rangle$, $[E_{12}]_{\mathcal{B}} = \langle 0, 1, 0, 0\rangle$, $[E_{21}]_{\mathcal{B}} = \langle 0, 0, 1, 0\rangle$, and $[E_{22}]_{\mathcal{B}} = \langle 0, 0, 0, 1\rangle$.

6. Can you make a conjecture about what the coordinate vectors should be for the basis elements of a basis in general?
   Answer: They should be the standard unit vectors $\vec{e}_i$ in R^k (where k is the number of vectors in the basis). This makes sense because a basis element is 1 times itself plus zero times each of the other basis elements.

Exercise 4.7.8 Let Z_s be the subset of $M_{2\times 2}$ whose entries sum to zero. That is,

$$Z_s = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \;\middle|\; a+b+c+d=0 \right\}.$$

Show that Z_s is a subspace of $M_{2\times 2}$.

Answer: The process followed in Example 4.7.2 can be followed here. The zero matrix is obviously an element of Z_s, so it's not empty. If we let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

be elements of Z_s, then $a_{11}+a_{12}+a_{21}+a_{22}=0$ as well as $b_{11}+b_{12}+b_{21}+b_{22}=0$. The sum

$$A+B = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{bmatrix},$$

and if we sum the elements, we get

$$a_{11}+b_{11}+a_{12}+b_{12}+a_{21}+b_{21}+a_{22}+b_{22} =$$
$$(a_{11}+a_{12}+a_{21}+a_{22})+(b_{11}+b_{12}+b_{21}+b_{22}) = 0$$

So $A+B$ is in Z_s, making it closed under vector addition. Similarly, for any scalar c,

$$cA = \begin{bmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{bmatrix},$$

and the sum of its entries is

$$ca_{11}+ca_{12}+ca_{21}+ca_{22} = c(a_{11}+a_{12}+a_{21}+a_{22}) = c(0) = 0.$$

Hence cA is in Z_s which is closed under scalar multiplication. We have demonstrated that Z_s is a subspace of $M_{2\times 2}$.

Alternatively, we can produce a spanning set. An example is

$$\left\{ \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

Exercise 4.7.9 Let N_s be the subset of $M_{2\times 2}$ whose entries sum to one. That is,

$$N_s = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \;\middle|\; a+b+c+d=1. \right\}.$$

Show that N_s is not a subspace of $M_{2\times 2}$.
Answer: We need only to show that one of the properties of a subspace is violated, and this can be done by producing one example showing that the set is not closed under one or the other operation (vector addition or scalar multiplication). There are lots of examples to choose from. A simple one is to consider the matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, which is clearly an element of N_s. If we take any scalar different from one, for example let $c = 2$, we find that

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

The entries of cA sum to two, not one. So cA is not an element of N_s. Since N_s is not closed under scalar multiplication, it is not a subspace of $M_{2\times 2}$. It is easy to verify that N_s is not closed under vector addition either.

Exercise 4.7.10 Let $\mathcal{D}$ be the subspace of $M_{2\times 2}$ of diagonal matrices. That is

$$\mathcal{D} = \left\{ \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \;\middle|\; a, d \in R \right\}.$$

Show that the set $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ is a basis for $\mathcal{D}$.
Answer: We have to show that $\mathcal{B}$ is linearly independent and spans $\mathcal{D}$. If we take any element of $\mathcal{D}$, we see that

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

So $\mathcal{B}$ spans $\mathcal{D}$. If we consider the homogeneous equation

$$c_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

then equating each entry, we see that $c_1 = c_2 = 0$. So $\mathcal{B}$ is linearly independent, and hence is a basis for $\mathcal{D}$.

Exercise 4.7.11: Let $S = \left\{ \vec{a}, \vec{b} \right\}$ where $a_n = 3$ for all n and $b_n = (-1)^n n^2$ for all n. Thus

$$\begin{aligned} \vec{a} &= \langle 3, 3, 3, 3, 3, \ldots \rangle \\ \vec{b} &= \langle -1, 4, -9, 16, -25, \ldots \rangle . \end{aligned}$$

1. Explain why the vector $\vec{0} = \langle 0,0,0,0,0,\ldots\rangle$ is in Span $\left\{\vec{a},\vec{b}\right\}$.

   Explanation:

$$\begin{aligned} 0\vec{a} + 0\vec{b} &= 0\langle 3,3,3,3,3,\ldots\rangle + 0\langle -1,4,-9,16,-25,\ldots\rangle \\ &= \langle 0,0,0,0,0,\ldots\rangle + \langle 0,0,0,0,0,\ldots\rangle \\ &= \langle 0,0,0,0,0,\ldots\rangle \\ &= \vec{0} \end{aligned}$$

   and thus $\vec{0}$ is a linear combination of $\vec{a}$ and $\vec{b}$. Therefore $\vec{0} \in$ Span $\left\{\vec{a},\vec{b}\right\}$.

3. Prove that Span $\left\{\vec{a},\vec{b}\right\} \neq R^{\infty}$.

   Proof: Span $\left\{\vec{a},\vec{b}\right\}$ consists of all linear combinations of $\vec{a}$ and $\vec{b}$. Hence every vector in Span $\left\{\vec{a},\vec{b}\right\}$ has the form

$$\begin{aligned} s\vec{a} + t\vec{b} &= s\langle 3,3,3,3,3,\ldots\rangle + t\langle -1,4,-9,16,-25,\ldots\rangle \\ &= \langle 3s-t, 3s+4t, 3s-9t, 3s+16t, 3s-25t, \ldots\rangle . \end{aligned}$$

   To show that Span $\left\{\vec{a},\vec{b}\right\} \neq R^{\infty}$, we just need to come up with an example of a vector, $\vec{c}$, in R^{∞} that is not a linear combination of $\vec{a}$ and $\vec{b}$. How do we make up such and example? We, let's see if we can come up with a $\vec{c}$ that has 1 and 11 as its first two components. Thus let us take $\vec{c}$ to be of the form

$$\vec{c} = \langle 1, 11, c_3, c_4, \ldots\rangle .$$

   In order to have $s\vec{a} + t\vec{b} = \vec{c}$ for some scalars s and t, we must have

$$\begin{aligned} 3s - t &= 1 \\ 3s + 4t &= 11. \end{aligned}$$

   The unique solution of this system of equations is $s = 1$, $t = 2$. This completely determines what every entry of $\vec{c}$ must be is we are to have $s\vec{a}+t\vec{b} = \vec{c}$. We must have $c_3 = 3(1)-9(2) = -15$, $c_4 = 3(1)+16(2) = 35$, etc. Hence the vector

$$\vec{c} = \langle 1, 11, 27, \text{anything}, \text{anything}, \ldots\rangle$$

   is not in Span $\left\{\vec{a},\vec{b}\right\}$.

Exercise 4.7.13

1. Let $\vec{a} = \langle 1, 2, 3, 4, 5, \ldots \rangle$ and $\vec{b} = \langle 2, 3, 4, 5, 6, \ldots \rangle$. What is the dimension of the subspace $S = \text{Span}\left\{\vec{a}, \vec{b}\right\}$? Explain.

   Answer: We claim that $\dim(S) = 2$. To prove this, we need to prove that S is linearly independent. This means we need to show that the equation $x_1\vec{a} + x_2\vec{b} = \vec{0}$ has only the trivial solution. Writing this equation out, we have

   $$x_1 \langle 1, 2, 3, 4, \ldots \rangle + x_2 \langle 2, 3, 4, 5, \ldots \rangle = \langle 0, 0, 0, 0, 0, \ldots \rangle$$

   or

   $$\langle x_1 + 2x_2, 2x_1 + 3x_2, 3x_1 + 4x_2, 4x_1 + 5x_2, \ldots \rangle = \langle 0, 0, 0, 0, 0, \ldots \rangle .$$

   This gives a system of linear equations with infinitely many equations and two unknowns:

   $$\begin{aligned} x_1 + 2x_2 &= 0 \\ 2x_1 + 3x_2 &= 0 \\ 3x_1 + 4x_2 &= 0 \\ &\text{etc.} \end{aligned}$$

   We only need to consider the first two equation in this system to prove our point. The first two equations are the system give the system

   $$\begin{aligned} x_1 + 2x_2 &= 0 \\ 2x_1 + 3x_2 &= 0 \end{aligned}$$

   which can easily be seen to have only the trivial solution $x_1 = x_2 = 0$. This means that the infinite set of equations also have only the trivial solution. Therefore S is linearly independent and $\dim(S) = 2$.

3. Let $\vec{e}_1 = \langle 1, 0, 0, 0, 0, \ldots \rangle$, $\vec{e}_2 = \langle 0, 1, 0, 0, 0, \ldots \rangle$ and $\vec{e}_3 = \langle 0, 0, 1, 0, 0, \ldots \rangle$. (Thus $\vec{e}_1$ has 1 as its first entry and all other entries are 0, etc.) What is the dimension of $\text{Span}\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$? Explain.

   Answer: We claim that $\dim(\text{Span}\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}) = 3$. To see why this is true, we need to show that the set $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ is linearly independent. Consider the equation

   $$x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3 = \vec{0}.$$

This is

$$\langle x_1, x_2, x_3, 0, 0, \ldots\rangle = \langle 0, 0, 0, 0, 0, \ldots\rangle$$

which clearly has only the trivial solution $x_1 = x_2 = x_3 = 0$.

Since $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ is a set of three linearly independent vectors, then $\dim(\mathrm{Span}\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}) = 3$.

Exercise 4.7.17 Let F_0 be the subset of $C^0(R)$ of functions that take the value of zero at zero. That is,

$$F_0 = \left\{ f \in C^0(R) \mid f(0) = 0 \right\}.$$

Determine whether F_0 is a subspace of $C^0(R)$. That is, either show that F_0 is a subspace of $C^0(R)$, or demonstrate that F_0 is not closed under vector addition or scalar multiplication.

Answer: F_0 is a subspace. First, the zero function $z(x) = 0$ is in F_0 because $z(0) = 0$. So F_0 is not empty. (We could use a more interesting example such as the sine function, since $\sin(x)$ is continuous on R and $\sin(0) = 0$.) If we consider any elements, f and g in F_0, then

$$f(0) = 0 \quad \text{and} \quad g(0) = 0.$$

Note that their sum, $f(x) + g(x)$ satisfies

$$f(0) + g(0) = 0 + 0 = 0.$$

And if c is any scalar, then the scalar multiple cf satisfies

$$cf(0) = c(0) = 0.$$

Hence F_0 is closed under both operations making it a subspace of $C^0(R)$

Exercise 4.7.20 Consider the vector space $\mathbb{P}_4$.

1. Determine which of the following are vectors in $\mathbb{P}_4$.

   (a) $p(x) = 2 + 3x - x^2 + 2x^3 + 4x^4$
   Answer This is a vector in $\mathbb{P}_4$.

   (b) $q(x) = 2 + 3x^2 - 9x^3 + 2x^4$
   Answer This is a vector in $\mathbb{P}_4$.

(c) $f(x) = -12 + x + 5x^2 - 6x^3$
Answer This is a vector in $\mathbb{P}_4$.

(d) $r(x) = 21x^3 - 4x^5$
Answer This is not a vector in $\mathbb{P}_4$. The degree exceeds 4.

2. Let $f(x) = 2x + x^3 - 14x^4$ and $g(x) = -3 + 4x^2 - 5x^3 + 10x^4$.

   (a) evaluate $2f(x)$
   Answer $2f(x) = 4x + 2x^3 - 28x^4$.

   (b) evaluate $3g(x)$
   Answer $3g(x) = -9 + 12x^2 - 15x^3 + 30x^4$.

   (c) evaluate $f(x) - g(x)$
   Answer $f(x) - g(x) = 3 + 2x - 4x^2 + 6x^3 - 24x^4$.

   (d) identify $-g(x)$
   Answer $-g(x) = 3 - 4x^2 + 5x^3 - 10x^4$

Exercise 4.7.21 Let $\mathcal{P}_{2,1}$ denote the set of all polynomials, $p(x) = p_0 + p_1x + p_2x^2$ of degree at most 2 with real coefficients that satisfy $p(1) = 0$.

1. Determine which of the following are elements of $\mathcal{P}_{2,1}$.

   (a) $g(x) = 2 - 3x - x^2$
   Answer: $g(1) = -2$, g is not in $\mathcal{P}_{2,1}$

   (b) $f(x) = 2 - 3x + x^2$
   Answer: $f(1) = 0$, f is in $\mathcal{P}_{2,1}$

   (c) $q(x) = 4x^2 + 2x - 6$
   Answer: $q(1) = 0$, q is in $\mathcal{P}_{2,1}$

2. Show that $\mathcal{P}_{2,1}$ is closed with respect to vector addition and scalar multiplication.
   Answer: Suppose f_1 and f_2 are elements of $\mathcal{P}_{2,1}$, and let c be any scalar. Note that since f_1 and f_2 are in $\mathcal{P}_{2,1}$, we know that $f_1(1) = 0$ and $f_2(1) = 0$. Using the definition of vector addition in $\mathbb{P}_2$, evaluating the sum at $x = 1$,

$$(f_1 + f_2)(1) = f_1(1) + f_2(1) = 0 + 0 = 0.$$

Similarly, the evaluating the scalar multiple cf_1 at $x = 1$,

$$(cf_1)(1) = cf_1(1) = c(0) = 0.$$

That is, the sum of elements in $\mathcal{P}_{2,1}$ are in $\mathcal{P}_{2,1}$, and scalar multiples of elements of $\mathcal{P}_{2,1}$ are in $\mathcal{P}_{2,1}$. This set is closed under vector addition and scalar multiplication.

3. Verify that every element of $\mathcal{P}_{2,1}$ can be written in the form $p(x) = p_1(x-1) + p_2(x^2-1)$. Note that we can say that $\mathcal{P}_{2,1} = \text{Span}\{x - 1, x^2 - 1\}$.
Answer: Let $p(x) = p_0 + p_1 x + p_2 x^2$ be any element of $\mathcal{P}_{2,1}$. Then

$$p(1) = p_0 + p_1(1) + p_2(1)^2 = p_0 + p_1 + p_2 = 0.$$

This gives an equation relating the values of p_0, p_1, and p_2, namely

$$p_0 = -p_1 - p_2.$$

So we can replace p_0 in $p(x)$ and write

$$p(x) = -p_1 - p_2 + p_1 x + p_2 x^2,$$

which we can rearrange in the form

$$p(x) = p_1(x-1) + p_2(x^2-1).$$

Exercise 4.7.23 Determine whether the indicated set is linearly independent or linearly dependent in the indicated vector space.

2. $\{1 + x, 1 - x\}$ in $\mathbb{P}_1$.
Answer: These are linearly independent.

3. $\{1 + x, 1 - x, 2 - 3x\}$ in $\mathbb{P}_1$.
Answer: These are linearly dependent. For example,

$$(1 + x) - 5(1 - x) + 2(2 - 3x) = 0 + 0x = \vec{0}(x).$$

Exercise 4.7.25: Show that the set of functions $S = \{e^x, e^{2x}\}$ is linearly independent and thus $\dim(\text{Span}(S)) = 2$.

Solution: We need to show that the equation

$$c_1 e^x + c_2 e^{2x} = 0 \text{ for all } x \in R$$

has only the trivial solution $c_1 = c_2 = 0$.

Since we require that the above equation is true for all $x \in R$, then we can set $x = 0$ to obtain

$$c_1 + c_2 = 0.$$

We can also set $x = \ln(2)$

$$c_1 e^{\ln(2)} + c_2 e^{\ln(4)} = 0$$

which can be written as

$$2c_1 + 4c_2 = 0.$$

The system of equations

$$\begin{aligned} c_1 + c_2 &= 0 \\ 2c_1 + 4c_2 &= 0 \end{aligned}$$

has only the trivial solution. Therefore S is linearly independent.

Exercise 4.8.2 Let $\mathcal{B} = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ be the ordered basis of $M_{2\times 2}$ where

$$E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Use coordinate vectors to determine whether the following collection of vectors is linearly dependent or linearly independent in $M_{2\times 2}$.

$$A_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 2 & -1 \\ 1 & -2 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 2 \\ 1 & 3 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 1 & 6 \\ 0 & 8 \end{bmatrix}.$$

Answer: We determine the coordinate vectors,

$$[A_1]_{\mathcal{B}} = \langle 1, 1, 0, 1 \rangle, \quad [A_2]_{\mathcal{B}} = \langle 2, -1, 1, -2 \rangle$$

$$[A_3]_{\mathcal{B}} = \langle 0, 2, 1, 3 \rangle, \quad [A_4]_{\mathcal{B}} = \langle 1, 6, 0, 8 \rangle.$$

Using these a column vectors, we can find an rref.

$$\begin{bmatrix} 1 & 2 & 0 & 1 \\ 1 & -1 & 2 & 6 \\ 0 & 1 & 1 & 0 \\ 1 & -2 & 3 & 8 \end{bmatrix} \xrightarrow{rref} \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We see that there is a non-pivot column, so the columns are linearly dependent. We can conclude that the set $\{A_1, A_2, A_3, A_4\}$ is linearly dependent in $M_{2\times 2}$.

Chapter 4 Additional Exercises

1. Prove statement a. of Theorem 4.1.2. That is, show that any set of vectors in R^n that includes the zero vector, $\vec{0}_n$, is linearly dependent.

   Answer: Let $k \geq 0$ and suppose we have a set of vectors $S = \left\{\vec{0}_n, \vec{u}_2, \vec{u}_3, \ldots, \vec{u}_k\right\}$ that includes the zero vector. Note that we can form the linear dependence relation

   $$1\vec{0}_n + 0\vec{u}_2 + 0\vec{u}_3 + \cdots + 0\vec{u}_k = \vec{0}_n.$$

   This is a valid linear dependence relation because at least one coefficient (the first one shown in the equation) is nonzero. Hence S is linearly dependent.

2. Prove that the set $\{\vec{0}_n\}$ is a subspace of R^n for any $n \geq 2$.

   Answer: The set if nonempty since is contains the zero vector. This vector also satisfies that all linear combinations of it are the zero vector (that is, when we scale the zero vector or add it to itself, we always get the zero vector). So the set is closed under both vector addition and scalar multiplication. So the set $\{\vec{0}_n\}$ satisfies the three conditions necessary to be a subspace.

3. Prove that if S is an subspace of R^n, then S must contain the zero vector, $\vec{0}_n$.

   Answer: Since S is nonempty (that's a property of a subspace), it contains at least one vector which we can call $\vec{x}$. As a subspace, S must be closed under scalar multiplication, so S must contain $c\vec{x}$ for every possible scalar c, including zero. So $0\vec{x} = \vec{0}_n$ must be an element of S.

4. Consider the subspace

$$S = \text{Span}\{\langle 1,2,1,1\rangle, \langle 3,5,3,4\rangle, \langle 1,1,1,2\rangle, \langle 2,1,1,4\rangle\}$$

of R^4. Find a basis for S.

Answer: If we create a matrix A having the vectors in the spanning set as its column vectors, then S will be its column space. A basis will be the pivot columns.

$$A = \begin{bmatrix} 1 & 3 & 1 & 2 \\ 2 & 5 & 1 & 1 \\ 1 & 3 & 1 & 1 \\ 1 & 4 & 2 & 4 \end{bmatrix} \quad \xrightarrow{rref} \quad \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The first, second and fourth columns are pivot columns, so a basis for S is

$$\{\langle 1,2,1,1\rangle, \langle 3,5,3,4\rangle, \langle 2,1,1,4\rangle\}.$$

5. Let $n \geq 2$; suppose S is a subspace of R^n and $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_k\}$ is a basis for S where $1 \leq k \leq n$. Explain why

$$[\vec{0}_n]_{\mathcal{B}} = \vec{0}_k.$$

That is, explain why the coordinate vector for the zero vector in S (which is the zero vector in R^n) must be the zero vector in R^k.

Answer: The elements of the basis are linearly independent. So the only coefficients that satisfy

$$\vec{0}_n = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k$$

are $c_1 = c_2 = \cdots = c_k = 0$. These are the entries of the coordinate vector $[\vec{0}_n]_{\mathcal{B}}$ making it the zero vector in R^k.

6. Determine whether the columns of A are linearly independent or linearly dependent. If the columns are linearly dependent, find a linear dependence relation.

Note: Each of these can be determined by setting up the homogeneous equation $A\vec{x} = \vec{0}_m$ (where m is the number of rows of A). If there are any non-pivot columns, they are linearly dependent, and a linear dependence relation is any nontrivial solution.

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

Answer: They're independent.

(b) $A = \begin{bmatrix} 1 & -2 & 0 & 1 \\ 2 & -4 & 2 & 6 \\ 0 & 0 & 3 & 6 \\ -3 & 6 & 1 & -1 \end{bmatrix}$

Answer: They are dependent. Any linear dependence relation will be some variation on

$$(2s - t)\,\mathrm{Col}_1(A) + s\,\mathrm{Col}_2(A) - 2t\,\mathrm{Col}_3(A) + t\,\mathrm{Col}_4(A) = \vec{0}_4.$$

(c) $A = \begin{bmatrix} 0 & 3 & 1 \\ 4 & 7 & 5 \\ -2 & -5 & -3 \\ 5 & -4 & 2 \end{bmatrix}$

Answer: They are dependent. Any linear dependence relation will be some variation on

$$2t\,\mathrm{Col}_1(A) + t\,\mathrm{Col}_2(A) + 3t\,\mathrm{Col}_3(A) = \vec{0}_4.$$

(d) $A = \begin{bmatrix} 3 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & -2 & 5 \\ -1 & 8 & 4 \end{bmatrix}$

Answer: These are independent.

7. Verify that $\mathbb{P}_n$ satisfies axioms 1–4 and axioms 7–10 of Definition 4.6.1. (Note that axioms 5 and 6 have already been discussed.)

Answer: The first two axioms 1 and 2 are the closure axioms. Suppose that p and q are elements of $\mathbb{P}_n$. This means that they are polynomials of degree at most n (i.e. n or less). Letting $p(x) = p_0 + p_1 x + \ldots + p_n x^n$ and $q(x) = q_0 + q_1 x + \ldots + q_n x^n$, then their sum

$$(p+q)(x) = (p_0+q_0) + (p_1+q_1)x + \ldots + (p_n+q_n)x^n,$$

and any scalar multiple of p,

$$cp(x) = cp_0 + cp_1 x + \ldots + cp_n x^n,$$

are also polynomials of degree at most n (the degree can decrease, but it can't increase). Hence axioms 1 and 2 are satisfied. To show that axiom 3 holds, we use the commutative property of real number (since the coefficients are real numbers). Note that

$$\begin{aligned} p(x)+q(x) &= (p_0+q_0) + (p_1+q_1)x + \ldots + (p_n+q_n)x^n \\ &= (q_0+p_0) + (q_1+p_1)x + \ldots + (q_n+p_n)x^n \\ &= q(x)+p(x). \end{aligned}$$

For axiom 4, let $r = r_0 + r_1 x + \ldots + r_n x^n$. Then

$$\begin{aligned} (p(x)+q(x))+r(x) &= (p_0+q_0) + (p_1+q_1)x + \ldots + (p_n+q_n)x^n + r_0 + r_1 x + \ldots + r_n x^n \\ &= (p_0+q_0+r_0) + (p_1+q_1+r_1)x + \ldots + (p_n+q_n+r_n)x^n \\ &= (p_0+(q_0+r_0)) + (p_1+(q_1+r_1))x + \ldots + (p_n+(q_n+r_n))x^n \\ &= p_0 + p_1 x + \ldots + p_n x^n + (q_0+r_0) + (q_1+r_1)x + \ldots + (q_n+r_n)x^n \\ &= p(x) + (q(x)) + r(x)). \end{aligned}$$

The additive identity and additive inverse properties have already been considered. Moving on to axiom 7, letting c be any scalar,

$$\begin{aligned} c\big(p(x)+q(x)\big) &= c\big((p_0+q_0) + (p_1+q_1)x + \ldots + (p_n+q_n)x^n\big) \\ &= c(p_0+q_0) + c(p_1+q_1)x + \ldots + c(p_n+q_n)x^n \\ &= (cp_0+cq_0) + (cp_1+cq_1)x + \ldots + (cp_n+cq_n)x^n \\ &= cp_0 + cp_1 x + \ldots + cp_n x^n + cq_0 + cq_1 x + \ldots + cq_n x^n \\ &= c(p_0 + p_1 x + \ldots + p_n x^n) + c(q_0 + q_1 x + \ldots + q_n x^n) \\ &= cp(x) + cq(x). \end{aligned}$$

Axiom 8 can be established similarly. Letting d be any scalar,

$$\begin{aligned}(c+d)p(x) &= (c+d)p_0 + (c+d)p_1x + \ldots + (c+d)p_nx^n \\ &= cp_0 + dp_0 + cp_1x + dp_1x + \ldots + cp_nx^n + dp_nx^n \\ &= cp_0 + cp_1x + \ldots + cp_nx^n + dp_0 + dp_1x + \ldots + dp_nx^n \\ &= cp(x) + dp(x).\end{aligned}$$

For axiom 9, we have

$$\begin{aligned}c\big(dp(x)\big) &= c\big(dp_0 + dp_1x + \ldots + dp_nx^n\big) \\ &= cdp_0 + cdp_1x + \ldots + cdp_nx^n \\ &= (cd)\big(p_0 + p_1x + \ldots + p_nx^n\big) \\ &= (cd)p(x) \\ &= (dc)p(x) \\ &= (dc)\big(p_0 + p_1x + \ldots + p_nx^n\big) \\ &= dcp_0 + dcp_1x + \ldots + dcp_nx^n \\ &= d\big(cp_0 + cp_1x + \ldots + cp_nx^n\big) \\ &= d\big(cp(x)\big)\end{aligned}$$

Then finally, axiom 10 follows immediately from the fact that $1p_i = p_i$ for each real number p_i. That is,

$$1p(x) = 1p_0 + 1p_1x + \ldots + 1p_nx^n = p_0 + p_1x + \ldots + p_nx^n = p(x).$$

8. For each statement, indicate whether the statement is true or false. Give a brief explanation of reason for each conclusion.

   (a) If A is an $n \times n$ matrix, then $\mathcal{RS}(A) = \mathcal{CS}(A)$.

   Answer: This is false. The have the same dimension (which is always true), but consider $\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ The row space is $\text{Span}\{\langle 1,2\rangle\}$, but the column space is $\text{Span}\{\langle 1,0\rangle\}$.

   (b) If A is an $n \times n$ matrix, then $\dim\big(\mathcal{RS}(A)\big) = \dim\big(\mathcal{CS}(A)\big)$.

   Answer: This is true. The rank, dimension of the row space and dimension of the column space are all the same number. It doesn't matter if the matrix is square or not.

(c) If A is a 3×3 matrix and $\operatorname{rank}(A) = 3$, then the homogeneous equation $A\vec{x} = \vec{0}_3$ has only the trivial solution.

Answer: True. This indicates that the nullity $(A) = 3 - 3 = 0$. So the null space only includes the zero vector.

(d) The dimension of $\mathbb{P}_5$ is $\dim(\mathbb{P}_5) = 5$.

Answer: This is false. The dimension is 6. In general $\dim(\mathbb{P}_n) = n + 1$. The simplest basis is $\{1, x, x^2, \ldots, x^n\}$ which has $n + 1$ elements.

(e) If A is an $m \times n$ matrix with linearly dependent columns, then the equation $A\vec{x} = \vec{0}_m$ must have infinitely many solutions.

Answer: This is true. The system must be consistent, and A will have non-pivot column(s) meaning there will be free variable(s).

(f) If A is an $m \times n$ matrix with linearly dependent columns, then the equation $A\vec{x} = \vec{y}$ must have infinitely many solutions for any $\vec{y}$ in R^m.

Answer: This is false. Such a system could be inconsistent. It is the case that every consistent system will have infinitely many solutions, but there's no general guarantee that the system is consistent.

(g) An element of a vector space is called a vector.

Answer: This is true. This is the most general definition of what a vector is.

(h) If p is a vector in $\mathbb{P}_4$ and $\mathcal{B}$ is some basis of $\mathbb{P}_4$, then the coordinate vector $[p]_{\mathcal{B}}$ is a vector in R^5.

Answer: This is true. Since $\dim(\mathbb{P}_4) = 5$, the basis will have five elements, and the coordinate vectors will be real five-tuples.

(i) For matrix A, if the first three rows of $\operatorname{rref}(A)$ are nonzero, then the first three rows of A are linearly independent.

Answer: This is false. Row operations allow for swapping the order of the rows and do not preserve the linear dependence relations of the rows.

(j) For matrix A, if the first three column vectors of $\operatorname{rref}(A)$ are three different standard unit vectors, then the first three columns of A

are linearly independent.

Answer: This is true. This would mean that the first three columns are pivot columns making them linearly independent.

9. For each matrix, find bases for the row space, the column space, and the null space.

(a) $A = \begin{bmatrix} 1 & 2 & -1 & 3 \\ 2 & 4 & -2 & 6 \\ 1 & 4 & 0 & 0 \end{bmatrix}$

Answer:

- A basis for $\mathcal{N}(A)$ is $\big\{\langle 2, -1/2, 1, 0\rangle, \langle -6, 3/2, 0, 1\rangle\big\}$.
- A basis for $\mathcal{CS}(A)$ is $\big\{\langle 1, 2, 1\rangle, \langle 2, 4, 4\rangle\big\}$.
- A basis for $\mathcal{RS}(A)$ is $\big\{\langle 1, 0, -2, 6\rangle, \langle 0, 1, 1/2, -3/2\rangle\big\}$.

(b) $B = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 5 & 0 \\ 4 & 6 & -2 \\ 2 & 3 & -1 \end{bmatrix}$

Answer:

- A basis for $\mathcal{N}(A)$ is $\big\{\langle 5, -3, 1\rangle\big\}$.
- A basis for $\mathcal{CS}(A)$ is $\big\{\langle 1, 3, 4, 2\rangle, \langle 2, 5, 6, 3\rangle\big\}$.
- A basis for $\mathcal{RS}(A)$ is $\big\{\langle 1, 0, -5\rangle, \langle 0, 1, 3\rangle\big\}$.

(c) $C = \begin{bmatrix} -2 & 2 & -3 & -2 & -8 \\ 3 & -3 & 3 & 1 & 10 \\ 2 & -2 & 2 & 0 & 4 \end{bmatrix}$

Answer:

- A basis for $\mathcal{N}(A)$ is $\big\{\langle 1, 1, 0, 0, 0\rangle, \langle -6, 0, 4, -4, 1\rangle\big\}$.
- A basis for $\mathcal{CS}(A)$ is $\big\{\langle -2, 3, 2\rangle, \langle -3, 3, 2\rangle, \langle -2, 1, 0\rangle\big\}$.
- A basis for $\mathcal{RS}(A)$ is $\big\{\langle 1, -1, 0, 0, 6\rangle, \langle 0, 0, 1, 0, -4\rangle, \langle 0, 0, 0, 1, 4\rangle\big\}$.

10. The first three *Chebyshev polynomials* are

$$T_0(x) = 1, \quad T_1(x) = x, \quad \text{and} \quad T_2(x) = 2x^2 - 1.$$

(It is customary to use the notation T_i when writing these polynomials.) Show that the set $\mathcal{C} = \{T_0, T_1, T_2\}$ is a basis for $\mathbb{P}_2$.

Answer: We have to show that $\mathcal{C}$ spans $\mathbb{P}_2$ and is linearly independent. If $p(x) = p_0 + p_1 x + p_2 x^2$ is any element of $\mathbb{P}_2$, we can write

$$p(x) = c_1 T_0(x) + c_2 T_1(x) + c_3 T_2(x),$$

where $c_1 = p_0 + \frac{1}{2}p_2$, $c_2 = p_1$, and $c_3 = \frac{1}{2}p_2$. To verify, note that

$$\begin{aligned}
c_1 T_0(x) + c_2 T_1(x) + c_3 T_2(x) &= c_1(1) + c_2(x) + c_3(2x^2 - 1) \\
&= (p_0 + \frac{1}{2}p_2)(1) + p_1 x + \left(\frac{1}{2}p_2\right)(2x^2 - 1) \\
&= p_0 + \frac{1}{2}p_2 + p_1 x + \frac{1}{2}p_2(2x^2) - \frac{1}{2}p_2 \\
&= p_0 + p_1 x + p_2 x^2 \\
&= p(x).
\end{aligned}$$

So $\mathcal{C}$ spans $\mathbb{P}_2$. To show that $\mathcal{C}$ is linearly independent, we can consider the homogeneous equation $c_1 T_0(x) + c_2 T_1(x) + c_3 T_2(x) = z(x)$.

$$\begin{aligned}
c_1 + c_2 x + c_3(2x^2 - 1) &= 0 + 0x + 0x^2 \\
(c_1 - c_3) + c_2 x + 2c_3 x^2 &= 0 + 0x + 0x^2.
\end{aligned}$$

Equating coefficients, we see that $c_1 - c_3 = 0$, $c_2 = 0$, and $2c_3 = 0$. The last two equations show that $c_2 = c_3 = 0$, and substituting $c_3 = 0$ into the first equation, we see that $c_1 = 0$ as well. The only solution is the trivial solution, so $\mathcal{C}$ is linearly independent. We can conclude that $\mathcal{C}$ is a basis for $\mathbb{P}_2$.

11. Suppose A is a 7×10 matrix.

    (a) If $\mathcal{RS}(A)$ is a subspace of R^k, what is k?

    Answer: $k = 10$ because the row vectors are vectors in R^{10}.

    (b) If $\mathcal{CS}(A)$ is a subspace of R^k, what is k?

    Answer: $k = 7$ because the column vectors are vectors in R^7.

    (c) If $\mathcal{N}(A)$ is a subspace of R^k, what is k?

    Answer: $k = 10$ because $A\vec{x}$ is defined for vectors in R^{10}.

(d) If $\text{rank}(A) = 7$, find $\text{nullity}(A)$.

Answer: $\text{nullity}(A) = 3$ because $\text{rank}(A) + \text{nullity}(A) = 10$, the total number of columns.

(e) If the homogeneous equation $A\vec{x} = \vec{0}_7$ has four free variables, what is $\dim(\mathcal{RS}(A))$?

Answer: $\dim((\mathcal{RS}(A)) = 6$. The given means that $\text{nullity}(A) = 4$. Since $\text{rank}(A) + \text{nullity}(A) = 10$, this would make $\text{rank}(A) = 6$, and the row space dimension is the same as the rank.

(f) If A is full rank, what is $\text{rank}(A)$? (Recall that full rank means that the rank is the largest it can be.)

Answer: $\text{rank}(A) = 7$. The largest the rank can be is the smaller of the number of rows and columns.

(g) If $\text{rank}(A) = 4$, find $\text{nullity}(A^T)$.

Answer: $\text{nullity}(A^T) = 3$. Since $\text{rank}(A^T) = \text{rank}(A)$, we can use the rank-nullity theorem. The n value for A^T is 7.

(h) If $\text{rank}(A^T) = 6$, what is $\dim(\mathcal{CS}(A))$?

Answer: $\dim(\mathcal{CS}(A)) = 6$ because $\text{rank}(A) = \text{rank}(A^T)$, and the rank is the dimension of the column space.

12. Which of the following sets, S, are subspaces of R^∞? Explain your answers.

(a) $S = \text{Span}\,\{\langle 1, 3, 5, 7, \ldots\rangle, \langle 2, 4, 6, 8, \ldots\rangle\}$

Answer: Yes. Any span of any set of vectors is a subspace.

(b) $S = \{\vec{a} = \langle a_1, a_2, a_3, \ldots\rangle \in R^\infty \mid a_n \geq 0 \text{ for all } n = 1, 2, 3, \ldots\}$

(c) $S = \{\vec{a} \in R^\infty \mid \vec{a} \text{ diverges}\}$ (*Note:* This question requires knowledge of Calculus II material.)

Answer: No, S is not a subspace of R^∞. It does not contain the zero vector $\vec{0} = \langle 0, 0, 0, \ldots\rangle$.

(d) $S = \{\vec{a} \in R^\infty \mid \text{all entries of } \vec{a} \text{ are either 0 or 1 or } -1\}$

(e) $S = \{\vec{a} \in R^\infty \mid \ \vec{a} \text{ has only finitely many non–zero entries}\}$

Answer: S is a subspace of R^∞. If we take any two vectors, $\vec{a}$ and $\vec{b}$ in S, then each of these vectors contains only finitely many non–zero entires. If we write these vectors out as

$$\vec{a} = \langle a_1, a_2, a_3, \ldots \rangle$$
$$\vec{b} = \langle b_1, b_2, b_3, \ldots \rangle,$$

then we know that there is some subscript M for which $a_n = 0$ for all $n > M$ and we know that there is some subscript N for which $b_n = 0$ for all $n > N$. This means that $a_n + b_n = 0$ for all $n > \max\{M, N\}$ and thus $\vec{a} + \vec{b} \in S$. In other words S is closed under vector addition. It is also easy to see that S is closed under scalar multiplication. Therefore S is a subspace of R^∞.

13. Let S be the set of all functions, f, in $C^1(R)$ that are equal to their derivative. In other words,

$$S = \left\{ f \in C^1(R) \mid f' = f \right\}.$$

   (a) Which of the following functions, with domain R, are in the set S?

      i. $f(x) = x$
      ii. $f(x) = x^2$
      iii. $f(x) = e^x$
      iv. $f(x) = 4e^x$
      v. $f(x) = 7$
      vi. $f(x) = \sin(x)$
      vii. $f(x) = e^{3x}$

   (b) Is S a subspace of $C^1(R)$? Explain.

14. Let T be the set of all functions, f, in $C^1(R)$ whose derivatives are equal to the function x^2. In other words,

$$S = \left\{ f \in C^1(R) \mid f'(x) = x^2 \text{ for all } x \in R \right\}.$$

   (a) Which of the following functions, with domain R, are in the set S?

i. $f(x) = x$ **Answer:** f is not in S because $f'(x) = 1$. ($f'(x) \neq x^2$.)

ii. $f(x) = x^2$

iii. $f(x) = x^3$ **Answer:** f is not in S.

iv. $f(x) = x^3 - 12$

v. $f(x) = \frac{1}{3}x^3 + 27$ **Answer:** f is in S.

vi. $f(x) = 3x^3$

(b) Is T a subspace of $C^1(R)$? Explain.

Answer: T is not a subspace of $C^1(R)$. There are many reasons that could be given. The simplest reason is that the zero vector of $C^1(R)$ is the function z defined by $z(x) = 0$ for all $x \in R$ and T does not contain this function because $z'(x) \neq x^2$. Any subspace, S, of any vector space, V, must contain the zero vector of V.

15. Let K be the set of all functions, f, in $C^0([-\pi, \pi])$ that satisfy

$$\int_{-\pi}^{\pi} f(x)\ dx = 0.$$

(a) Which of the following functions, with domain $[-\pi, \pi]$, are in the set K?

i. $f(x) = x$

ii. $f(x) = x^2$

iii. $f(x) = x^3$

iv. $f(x) = \sin(x)$

v. $f(x) = \cos(x)$

vi. $f(x) = x\sin(x)$

vii. $f(x) = x\cos(x)$

(b) Is K a subspace of $C^0([-\pi, \pi])$? Explain.

16. Let L be the set of all functions, f, in $C^0([-\pi, \pi])$ that satisfy

$$\int_{-\pi}^{\pi} f(x)\ dx = 1.$$

Is L a subspace of $C^0([-\pi, \pi])$? Explain.

Answer: L is not a subspace of of $C^0([-\pi, \pi])$. The zero vector of $C^0([-\pi, \pi])$ is the function z defined by $z(x) = 0$ for all $x \in [-\pi, \pi]$ and this function is not in L because

$$\int_{-\pi}^{\pi} z(x)\, dx = 0 \neq 1.$$

17. Consider the set of functions $S = \{1, e^x\}$ in $F(R)$. Show that this set of functions is linearly independent and is thus a basis for $\text{Span}(S)$. Determine $[7 - 8e^x]_S$.

    Solution: We must show that the equation

    $$c_1(1) + c_2 e^x = 0 \text{ for all } x \in R$$

    has only the trivial solution. Setting $x = 0$ and $x = \ln(2)$ we obtain

    $$\begin{aligned} c_1 + c_2 &= 0 \\ c_1 + 2c_2 &= 0. \end{aligned}$$

    This system has only the trivial solution, so $S = \{1, e^x\}$ is linearly independent. Since

    $$7 - 8e^x = 7(1) + (-8)\, e^x,$$

    then

    $$[7 - 8e^x]_S = \langle 7, -8 \rangle .$$

18. Consider the set

    $$V = \{\vec{x} = \langle x_1, x_2 \rangle \mid x_1 \in R \text{ and } x_2 \in R\}.$$

    In other words, $V = R^2$, but we are going to define one of the operations on V differently than how we defined it for R^2 in Chapter 1. Thus V and R^2 are equal as sets, but not as vector spaces.

    We will define addition of elements of V in the usual way: For any elements $\vec{x}$ and $\vec{y}$ in V we define

    $$\vec{x} + \vec{y} = \langle x_1, x_2 \rangle + \langle y_1, y_2 \rangle = \langle x_1 + y_1, x_2 + y_2 \rangle .$$

However, we will define scalar multiplication in a different way: For any element $\vec{x} \in V$ and scalar $c \in R$ we define

$$c\vec{x} = \langle 0, 0 \rangle .$$

Explain why V with the operations defined as we have defined them above is not a vector space. This is a good exercise in understanding Definition 4.6.1. (Which of the ten axioms of Definition 4.6.1 does V satisfy and which of the axioms does it not satisfy?)

A.5 Chapter 5 Exercises:

Exercise 5.1.3 Let $f : D \to R$ be defined by the formula

$$f(x) = \frac{1}{x}.$$

1. What is the largest possible subset of R that you can choose for the domain, D, such that this formula makes sense (meaning that the formula produces a real number for all $x \in D$)?

   Answer: $1/x$ is defined for all real numbers $x \neq 0$, so the largest possible set we can choose for the domain is $D = (-\infty, 0) \cup (0, \infty)$.

2. Draw the graph of f (either by hand or using technology).

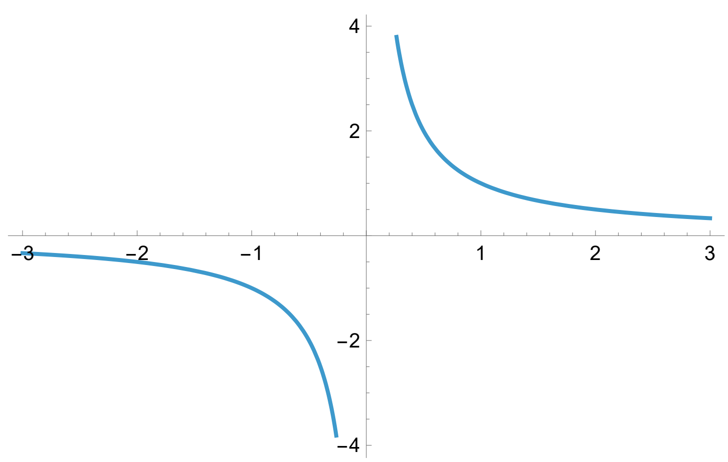


Figure A.4: Graph of $f(x) = 1/x$

3. Using the domain, D, that you designated in part 1, what is Range (f)?

   Answer: The range of f is Range $(f) = (-\infty, 0) \cup (0, \infty)$.

4. Do you see that f maps D onto its range and that f is one–to–one?

5. Find the formula for f^{-1}.

   Answer: $f^{-1}(x) = \frac{1}{x}$.

Exercise 5.1.4

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

Solution: The function T defined by $T(\vec{x}) = A\vec{x}$ maps R^2 into R^2, so the domain of T is R^2 and the codomain of T is R^2. Since

$$A = \begin{bmatrix} -3 & 1 \\ 1 & -2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \text{rref}(A),$$

we see that every row of of A contains a pivot which tells us that T maps R^2 **onto** R^2 and hence

$$\text{Range}(T) = \text{Span}\{\langle -3, 1\rangle, \langle 1, -2\rangle\} = R^2.$$

Since every column of A contains a pivot, then T is one–to–one. Since T is onto R^2 and one–to–one, then T is invertible. Also, $T^{-1}(\vec{x}) = A^{-1}\vec{x}$ were

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

The formula for T^{-1} is

$$T^{-1}(\langle x_1, x_2\rangle) = \left\langle -\frac{2}{5}x_1 - \frac{1}{5}x_2, -\frac{1}{5}x_1 - \frac{3}{5}x_2 \right\rangle.$$

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

Solution: The function T defined by $T(\vec{x}) = A\vec{x}$ maps R^2 into R^2, so the domain of T is R^2 and the codomain of T is R^2. Since

$$A = \begin{bmatrix} 0 & 3 \\ 0 & -4 \end{bmatrix} \to \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \text{rref}(A),$$

we see that not every row of of A contains a pivot which tells us that T does **not** map R^2 onto R^2. We see that

$$\text{Range}(T) = \text{Span}\{\langle 3, -4\rangle\}.$$

Since not every column of A contains a pivot, then T is **not** one–to–one. Since T is not onto R^2 and not one–to–one, then T is not invertible.

5.

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

Solution: The function T defined by $T(\vec{x}) = A\vec{x}$ maps R^3 into R^3, so the domain of T is R^3 and the codomain of T is R^3. Since

$$A = \begin{bmatrix} 3 & 0 & 1 \\ 1 & -4 & 1 \\ 4 & -4 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & -\frac{1}{6} \\ 0 & 0 & 0 \end{bmatrix} = \text{rref}(A),$$

we see that not every row of of A contains a pivot which tells us that T does not map R^3 onto R^3. We see that

$$\text{Range}(T) = \text{Span}\{\langle 3, 1, 4\rangle, \langle 0, -4, -4\rangle\}.$$

Since not every column of A contains a pivot, then T is not one–to–one. Thus T is not invertible.

7.

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

Solution: The function T defined by $T(\vec{x}) = A\vec{x}$ maps R^4 into R^3, so the domain of T is R^4 and the codomain of T is R^3. Since

$$A = \begin{bmatrix} -2 & -4 & 4 & 0 \\ 0 & 2 & 3 & -4 \\ 2 & -3 & 1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & -\frac{6}{31} \\ 0 & 1 & 0 & -\frac{23}{31} \\ 0 & 0 & 1 & -\frac{26}{31} \end{bmatrix} = \text{rref}(A),$$

we see that every row of of A contains a pivot which tells us that T maps R^4 onto R^3. We see that

$$\text{Range}\,(T) = R^3.$$

Since not every column of A contains a pivot, then T is not one–to–one. Since T is not one–to–one, then T is not invertible.

9.
$$A = \begin{bmatrix} 4 & 2 \end{bmatrix}$$

Solution: The function T defined by $T(\vec{x}) = A\vec{x}$ maps R^2 into R^1, so the domain of T is R^2 and the codomain of T is R^1. Since

$$A = \begin{bmatrix} 4 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & \frac{1}{2} \end{bmatrix} = \text{rref}\,(A)\,,$$

we see that every row of of A contains a pivot which tells us that T maps R^2 onto R^1. We see that

$$\text{Range}\,(T) = R^1.$$

Since not every column of A contains a pivot, then T is not one–to–one. Since T is not one–to–one, then T is not invertible.

Exercise 5.1.6 Suppose that A is an $m \times n$ matrix and suppose that $T : R^n \rightarrow R^m$ is the function defined by $T(\vec{x}) = A\vec{x}$. Explain why if $m \neq n$, then T is not invertible.

Answer: If $m > n$, then it is not possible that every row of A contains a pivot and thus it is not possible that T maps R^n onto R^m and thus it is not possible that T is invertible. If $m < n$, then it is not possible that every column of A contains a pivot and thus it is not possible that T is one–to–one and thus it is not possible that T is invertible. In conclusion, it is not possible for T to be invertible unless $m = n$.

Exercise 5.2.2 Determine whether or not each of the following expressions defines a linear transformation $T : R^n \rightarrow R^m$ (for appropriate m and n).

1. $T(\langle x_1, x_2 \rangle) = \langle -4x_1 - 4x_2, -5x_1 + 3x_2 \rangle$

   Answer: This is a linear transformation.

3. $T(\langle x_1, x_2 \rangle) = \langle x_1, 6x_2, -3x_1 \rangle$

   Answer: This is a linear transformation.

5. $T(\langle x_1, x_2, x_3 \rangle) = \left\langle \sqrt{x_1^2 + x_2^2 + x_3^2}, 0 \right\rangle$

   Answer: This is a not a linear transformation.

7. $T(\langle x_1, x_2 \rangle) = \langle 0, 0 \rangle$

   Answer: This is a linear transformation.

Exercise 5.2.4 Illustrate Theorem 5.2.3 for the linear transformations $T : R^n \to R^m$ given in 1-5.

1. $T(\langle x_1, x_2 \rangle) = \langle 3x_1 + 4x_2, 4x_2 \rangle$

   Solution: First note that $T : R^2 \to R^2$. The matrix for T is

   $$A = \begin{bmatrix} 3 & 4 \\ 0 & 4 \end{bmatrix}.$$

   Since

   $$A = \begin{bmatrix} 3 & 4 \\ 0 & 4 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \text{rref}(A),$$

   we see that

   $$\text{Range}(T) = R^2$$
   $$\ker(T) = \left\{ \vec{0}_2 \right\}.$$

   We also see that

   $$\dim(\text{Range}(T)) + \dim(\ker(T)) = 2 + 0 = 2.$$

3. $T(\langle x_1, x_2, x_3 \rangle) = \langle x_1 + x_2 + x_3, x_1 + x_2 + x_3, x_1 + x_2 + x_3 \rangle$

   Solution: First note that $T : R^3 \to R^3$. The matrix for T is

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

Since

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \text{rref}(A),$$

we see that

$$\begin{aligned} \text{Range}(T) &= \text{Span}\{\langle 1, 1, 1 \rangle\} \\ \ker(T) &= \text{Span}\{\langle -1, 1, 0 \rangle, \langle -1, 0, 1 \rangle\}. \end{aligned}$$

We have

$$\dim(\text{Range}(T)) + \dim(\ker(T)) = 1 + 2 = 3.$$

5. $T(\langle x_1, x_2, x_3 \rangle) = \langle 0, 0, 0, 0, 0 \rangle$.

Solution: First note that $T : R^3 \to R^5$. The matrix for T is

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

Since

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \text{rref}(A),$$

we see that

$$\begin{aligned} \text{Range}(T) &= \text{Span}\left\{\vec{0}_5\right\} \\ \ker(T) &= R^3. \end{aligned}$$

We have

$$\dim(\text{Range}(T)) + \dim(\ker(T)) = 0 + 3 = 3.$$

Exercise 5.2.6 Suppose that $T : R^4 \to R^5$ is a linear transformation. Explain why it is not possible that $\text{Range}\,(T) = R^5$.

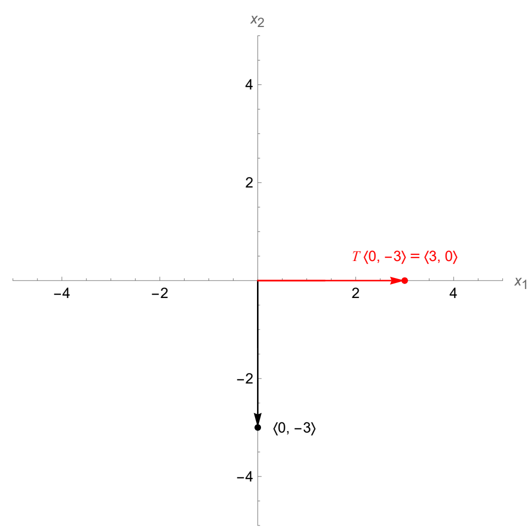
Explanation: If $\text{Range}\,(T) = R^5$, then $\dim\,(\text{Range}\,(T)) = 5$. However, by the Fundamental Theorem of Linear Algebra, we must have $\dim\,(\text{Range}\,(T)) + \dim\,(\ker\,(T)) = 4$. This would not be possible if $\dim\,(\text{Range}\,(T)) = 5$.

Exercise 5.2.8 Suppose that $Z : R^4 \to R^4$ is the **zero transformation**, which is defined by $Z\,(\vec{x}) = \vec{0}_4$ for all $\vec{x} \in R^4$. What are the dimensions of $\text{Range}\,(Z)$ and $\ker\,(Z)$?

Answer: $\dim\,(\text{Range}\,(Z)) = 0$ and $\dim\,(\ker\,(Z)) = 4$.

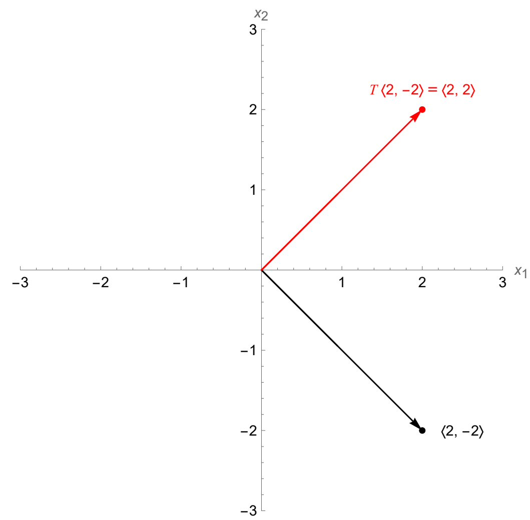
Exercise 5.3.1

1. We have $T\langle 0, -3\rangle = \langle 3, 0\rangle$. A picture of the input vector and output vector is shown below.



Exercise 5.3.3 For the vector $\vec{x} = \langle 1, -1\rangle$, we have

$$\begin{aligned} 2\vec{x} &= \langle 2, -2\rangle \\ T\,(\vec{x}) &= \langle 1, 1\rangle \\ T\,(2\vec{x}) &= \langle 2, 2\rangle \\ 2T\,(\vec{x}) &= \langle 2, 2\rangle \end{aligned}$$



Exercise 5.3.5

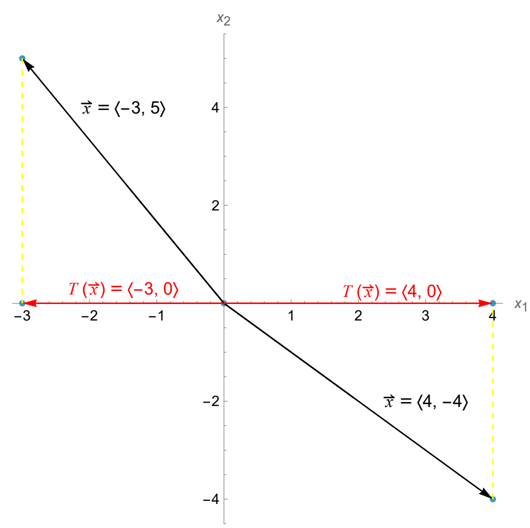


Figure A.5: $T(\langle x_1, x_2\rangle) = \langle x_1, 0\rangle$

Exercise 5.3.7 For the linear transformation $T(\langle x_1, x_2\rangle) = \langle x_1, 0\rangle$ we have

$$T(\langle 1, 0\rangle) = \langle 1, 0\rangle$$
$$T(\langle 0, 1\rangle) = \langle 0, 0\rangle$$

so the standard matrix for T is

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

A already has reduced echelon form and we see that

$$\begin{aligned} \text{Range}\,(T) &= \text{Span}\,(\langle 1, 0\rangle) \\ \ker\,(T) &= \text{Span}\,(\langle 0, 1\rangle)\,. \end{aligned}$$

T does not map R^2 onto R^2 and T is not one–to–one. Thus T is not invertible.

Exercise 5.3.10 For the linear transformation $T(\vec{x}) = 2\vec{x}$ we have

$$\begin{aligned} T(\langle 1, 0\rangle) &= 2\langle 1, 0\rangle = \langle 2, 0\rangle \\ T(\langle 0, 1\rangle) &= 2\langle 0, 1\rangle = \langle 0, 2\rangle \end{aligned}$$

and thus the matrix for T is

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

Since

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \text{rref}\,(A)\,,$$

we see that

$$\begin{aligned} \text{Range}\,(T) &= \text{Span}\,\{\langle 2, 0\rangle\,, \langle 0, 2\rangle\} = R^2 \\ \ker\,(T) &= \left\{\vec{0}_2\right\}. \end{aligned}$$

T maps R^2 onto R^2 and is one–to–one, so T is invertible. The matrix for T^{-1} is

$$A^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix},$$

so the formula for T^{-1} is

$$T^{-1}(\langle x_1, x_2\rangle) = \left\langle \frac{1}{2}x_1, \frac{1}{2}x_2 \right\rangle$$

or simply $T^{-1}(\vec{x}) = \frac{1}{2}\vec{x}$.

Exercise 5.3.11 This transformation is $T(\vec{x}) = -4\vec{x}$.

Exercise 5.3.14 Concerning the reflection transformation $T(\langle x_1, x_2\rangle) = \langle x_2, x_1\rangle$, the matrix for this transformation is

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

Since

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \text{rref}(A),$$

then

$$\text{Range}(T) = R^2$$
$$\ker(T) = \left\{\vec{0}_2\right\}.$$

T maps R^2 onto R^2 and is one–to–one, so it is invertible. T^{-1} has matrix

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

so $T^{-1}(\langle x_1, x_2\rangle) = \langle x_2, x_1\rangle$. Note that $T^{-1} = T$. This makes sense because to "undo" reflection through the line $x_2 = x_1$, we repeat the same action (reflect again).

Exercise 5.3.18 The shearing transformation $T(\langle x_1, x_2\rangle) = \langle x_1 + x_2, x_2\rangle$ has matrix

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

Since

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \text{rref}(A),$$

then

$$\text{Range}(T) = R^2$$
$$\ker(T) = \left\{\vec{0}_2\right\}.$$

T maps R^2 onto R^2 and is one–to-one. Therefore T is invertible. The matrix for T^{-1} is

$$A^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

and thus the formula for T^{-1} is $T^{-1}(\langle x_1, x_2 \rangle) = \langle x_1 - x_2, x_2 \rangle$.

Note that

$$\dim(\text{Range}(T)) + \dim(\ker(T)) = 2 + 0 = 2.$$

Exercise 5.3.20 Based on our knowledge of the unit circle, we see that

$$T(\langle 1, 0 \rangle) = \left\langle \frac{\sqrt{3}}{2}, \frac{1}{2} \right\rangle$$
$$T(\langle 0, 1 \rangle) = \left\langle -\frac{1}{2}, \frac{\sqrt{3}}{2} \right\rangle$$

so the matrix for R_{30° is

$$A_{30^\circ} = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}.$$

The matrix for $R_{30^\circ}^{-1} = R_{-30^\circ}$ is

$$(A_{30^\circ})^{-1} = A_{-30^\circ} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}.$$

Exercise 5.3.22 The matrix for A_θ is

$$A_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

The matrix for $(A_\theta)^{-1} = A_{-\theta}$ is

$$(A_\theta)^{-1} = A_{-\theta} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}.$$

We see that

$$\begin{aligned}
A_\theta^{-1}A_\theta &= \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \\
&= \begin{bmatrix} \cos^2(\theta) + \sin^2(\theta) & -\cos(\theta)\sin(\theta) + \sin(\theta)\cos(\theta) \\ -\sin(\theta)\cos(\theta) + \cos(\theta)\sin(\theta) & \cos^2(\theta) + \sin^2(\theta) \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&= I_2.
\end{aligned}$$

Exercise 5.3.24 If $\vec{x} = \langle x_1, x_2 \rangle$ is any vector in R^2 and θ is any angle then

$$\vec{x} \cdot R_\theta(\vec{x}) = \langle x_1, x_2 \rangle \cdot \langle \cos(\theta)x_1 - \sin(\theta)x_2, \sin(\theta)x_1 + \cos(\theta)x_2 \rangle$$

$$\begin{aligned}
&= x_1(\cos(\theta)x_1 - \sin(\theta)x_2) + x_2(\sin(\theta)x_1 + \cos(\theta)x_2) \\
&= \cos(\theta)x_1^2 - \sin(\theta)x_1x_2 + \sin(\theta)x_1x_2 + \cos(\theta)x_2^2 \\
&= \left(x_1^2 + x_2^2\right)\cos(\theta) \\
&= \|\vec{x}\|^2 \cos(\theta).
\end{aligned}$$

If $\vec{x} \neq \vec{0}_2$ and θ is an acute angle, then $\|\vec{x}\|^2 > 0$ and $\cos(\theta) > 0$ and hence $\vec{x} \cdot R_\theta(\vec{x}) > 0$.

If $\vec{x} \neq \vec{0}_2$ and θ is an obtuse angle, then $\|\vec{x}\|^2 > 0$ and $\cos(\theta) < 0$ and hence $\vec{x} \cdot R_\theta(\vec{x}) < 0$.

If $\vec{x} \neq \vec{0}_2$ and $\theta = 90°$, then $\|\vec{x}\|^2 > 0$ and $\cos(\theta) = 0$ and hence $\vec{x} \cdot R_\theta(\vec{x}) = 0$.

If $\vec{x} = \vec{0}_2$, then $\|\vec{x}\|^2 = 0$ and hence $\vec{x} \cdot R_\theta(\vec{x}) = 0$.

Exercise 5.4.1 Here are the solutions for numbers 1 and 3.

1) The line L that contains the points $P = (3, 1)$ and $Q = (2, -4)$ has slope

$$m = \frac{-4 - 1}{2 - 3} = 5$$

so L has equation

$$x_2 - 1 = 5(x_1 - 3)$$

which can be written as

$$x_2 = 5x_1 - 14.$$

3) The line L that contains the points $P = (3,5)$ and $Q = (-12,5)$ has slope

$$m = \frac{5-5}{-12-3} = 0.$$

so L has equation

$$x_2 - 5 = 0\,(x_1 - 3)$$

which can be written as

$$x_2 = 5.$$

Exercise 5.4.3 The points on L all have the form $(x_1, 2x_1 + 1)$.

When we apply T to such a point, then we get the point $(-2x_1 - 1, x_1)$. Thus, for the point that we get, the first coordinate is -2 times the second coordinate minus 1. This means that this point lies on the line

$$T\,(L) : x_1 = -2x_2 - 1$$

which can be written as

$$T\,(L) : x_2 = -\frac{1}{2}x_1 - \frac{1}{2}. \tag{A.5}$$

Another way to think about this problem is to write the equation for L in vector parametric form. Since L has slope 2, then a direction vector for L is $\vec{d} = \langle 1, 2\rangle$. In addition, L contains the point $P = (1,3)$. Hence a vector equation for L is

$$L : \vec{x} = \langle 1, 3\rangle + t\,\langle 1, 2\rangle\,.$$

Thus, for any point on L, we have

$$T\,(\vec{x}) = T\,(\langle 1,3\rangle) + tT\,(\langle 1,2\rangle) = \langle -3, 1\rangle + t\,\langle -2, 1\rangle\,.$$

Hence $T\,(L)$ is the line that contains the point $(-3,1)$ and has direction vector $\langle -2, 1\rangle$. (This means that L has slope $-1/2$.) This agrees with the equation we wrote for $T\,(L)$ in (A.5).

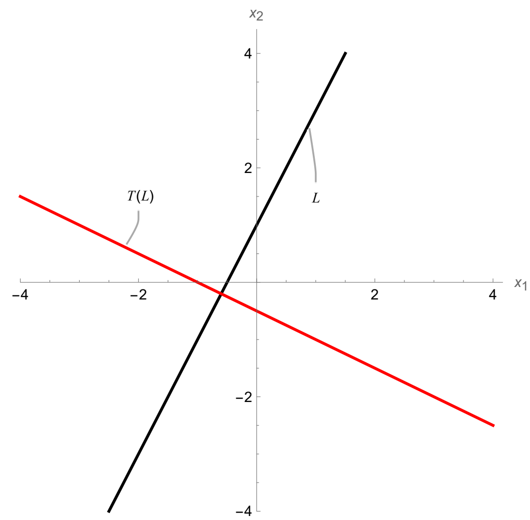


Figure A.6: L and $T(L)$ for Exercise 5.4.3

Exercise 5.4.5 This linear transformation maps all lines to lines. Its kernel is $\left\{\vec{0}_2\right\}$.

Exercise 5.4.7

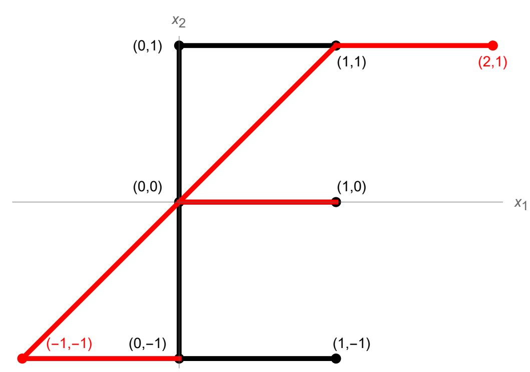


Figure A.7: Sheared E is in Red

Example 5.5.1

For the linear transformations

$$S(\langle x_1, x_2, x_3\rangle) = \langle 2x_1 + x_2, 2x_1 + x_2 + x_3\rangle$$
$$T(\langle x_1, x_2\rangle) = \langle -x_1, 3x_1 - x_2, -2x_1 + 3x_2\rangle :$$

1. $S \circ T$ is a linear transformation from R^2 to R^2.

2. The standard matrices for S and T are

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

3. The standard matrix for $S \circ T$ is

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

4. The formula for $S \circ T$ is

$$(S \circ T)(\langle x_1, x_2\rangle) = \langle x_1 - x_2, -x_1 + 2x_2\rangle .$$

Exercise 5.5.3In Section 5.3.6, we studied the rotation transformations R_θ. Let $R_{45^\circ} : R^2 \to R^2$ be the linear transformation that rotates vectors in R^2 counterclockwise through angle 45°.

1. What does the linear transformation $R_{45^\circ} \circ R_{45^\circ}$ do to vectors in R^2? Explain in words and fill in the blank below

   Answer: $R_{45^\circ} \circ R_{45^\circ}$ rotates a vector by 45° and then rotates by 45° again. Hence $R_{45^\circ} \circ R_{45^\circ}$ rotates vectors by 90°.

$$R_{45^\circ} \circ R_{45^\circ} = R_{90^\circ}.$$

2. Show that $A_{45^\circ} A_{45^\circ} = A_{90^\circ}$.

$$\begin{aligned} A_{45^\circ} A_{45^\circ} &= \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \\ &= \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \\ &= A_{90^\circ}. \end{aligned}$$

3. Without doing any computations, what do you guess that we get when we multiply the matrix A_{45° by itself eight times? In other words, what do you guess is $(A_{45^\circ})^8$. After guessing, compute $(A_{45^\circ})^8$ to see if your guess is correct.

   Answer: Composing R_{45° with itself eight times performs eight successive 45° counterclockwise rotations. This is the same as a 360° rotation, which is the same as doing nothing. Thus we guess that $(A_{45^\circ})^8 = I_2$. Using a calculator, we obtain

$$(A_{45^\circ})^8 = \left(\begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \right)^8 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Exercise 5.5.5

The matrix for R_{30° is

$$A_{30^\circ} = \begin{bmatrix} \cos(30^\circ) & -\sin(30^\circ) \\ \sin(30^\circ) & \cos(30^\circ) \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix}$$

and the matrix for $(R_{30^\circ})^{-1} = R_{-30^\circ}$ is

$$A_{-30^\circ} = \begin{bmatrix} \cos(-30^\circ) & -\sin(-30^\circ) \\ \sin(-30^\circ) & \cos(-30^\circ) \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 & 1/2 \\ -1/2 & \sqrt{3}/2 \end{bmatrix}.$$

We observe that

$$A_{30^\circ} A_{-30^\circ} = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix} \begin{bmatrix} \sqrt{3}/2 & 1/2 \\ -1/2 & \sqrt{3}/2 \end{bmatrix} = I_2.$$

Exercise 5.5.7

The matrix for R_{60° is

$$A_{60^\circ} = \begin{bmatrix} \cos(60^\circ) & -\sin(60^\circ) \\ \sin(60^\circ) & \cos(60^\circ) \end{bmatrix} = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}.$$

The matrix for S is

$$A_S = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The matrix for P is

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

In Example 5.5.3, we found that the matrix for $P \circ S \circ R_{60^\circ}$ is

$$A_P A_S A_{60^\circ} = \begin{bmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & -1 \end{bmatrix}$$

1. Is $P \circ R_{60^\circ} \circ S$ the same or different from $P \circ S \circ R_{60^\circ}$?

   Answer: The matrix for $P \circ R_{60^\circ} \circ S$ is

   $$\begin{aligned} A_P A_{60^\circ} A_S &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}. \end{aligned}$$

   Hence $P \circ R_{60^\circ} \circ S$ is different from $P \circ S \circ R_{60^\circ}$

2. Is $S \circ R_{60^\circ} \circ P$ the same or different from $P \circ S \circ R_{60^\circ}$?

   Answer: The matrix for $S \circ R_{60^\circ} \circ P$ is

   $$\begin{aligned} A_S A_{60^\circ} A_P &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & -1 \end{bmatrix}, \end{aligned}$$

   so $S \circ R_{60^\circ} \circ P$ the the same as $P \circ S \circ R_{60^\circ}$.

Exercise 5.5.9

1. Explain why any $n \times n$ matrix, A, is similar to itself.

   Explanation: Since $A = I_n^{-1} A I_n$, then A is similar to A.

2. Explain why if A is similar to B, then B is similar to A.

   Explanation: Suppose that A is similar to B. Then there exists an invertible $n \times n$ matrix C such that

   $$A = C^{-1} B C.$$

This gives
$$CA = C\left(C^{-1}BC\right)$$
which gives
$$CA = \left(CC^{-1}\right)(BC) = I_n(BC) = BC.$$
We now have that $BC = CA$. This gives
$$(BC)\,C^{-1} = (CA)\,C^{-1}$$
and hence
$$B = CAC^{-1}$$
or equivalently
$$B = \left(C^{-1}\right)^{-1} AC^{-1}.$$
This shows that B is similar to A.

3. Explain why if A is similar to B and B is similar to C, then A is similar to C.

   Explanation: Suppose that A is similar to B and B is similar to C. Then there exist invertible $n \times n$ matrices D and E such that
$$A = D^{-1}BD \text{ and } B = E^{-1}CE.$$
   This gives
$$\begin{aligned} A &= D^{-1}BD \\ &= D^{-1}\left(E^{-1}CE\right)D \\ &= \left(D^{-1}E^{-1}\right)C\,(ED) \\ &= (ED)^{-1}\,C\,(ED)\,, \end{aligned}$$
   which shows that A is similar to C.

Exercise 5.5.11 Suppose that $T : R^2 \to R^2$ is a linear transformation that is similar to the identity transformation $E(\vec{x}) = \vec{x}$. Then there exists an invertible linear transformation $P : R^2 \to R^2$ such that $T = P^{-1} \circ E \circ P$. However, the identity transformation does nothing to vectors in R^2. This means that composing any transformation with the identity transformation just gives the original transformation. Thus

$$\begin{aligned} T &= P^{-1} \circ E \circ P \\ &= P^{-1} \circ (E \circ P) \\ &= P^{-1} \circ P \\ &= E. \end{aligned}$$

Exercise 5.6.1
Proof: Suppose that $\vec{a} = \langle a_1, a_2, a_3, \ldots \rangle \in R^\infty$ and that c is a scalar. Then

$$\begin{aligned} S(c\vec{a}) &= S(c\langle a_1, a_2, a_3, \ldots \rangle) \\ &= S(\langle ca_1, ca_2, ca_3, \ldots \rangle) \\ &= \langle ca_2, ca_3, ca_4, \ldots \rangle \\ &= c\langle a_2, a_3, a_4, \ldots \rangle \\ &= cS(\vec{a}) \end{aligned}$$

Exercise 5.6.4
Proof: Suppose that $\vec{x} \in W$ and suppose that c is a scalar. Since T is invertible then there is a unique vector $\vec{u} \in V$ such that $T(\vec{u}) = \vec{x}$ and hence $\vec{u} = T^{-1}(\vec{x})$. Since T is a linear transformation, then

$$c\vec{x} = cT(\vec{u}) = T(c\vec{u})$$

and thus

$$T^{-1}(c\vec{x}) = c\vec{u} = cT^{-1}(\vec{x}).$$

This shows that the second requirement of Definition 5.6.1 is satisfied.

Exercise 5.6.6 **Proof:** We need to show that

$$\ker(T) = \left\{ \vec{x} \in V \mid T(\vec{x}) = \vec{0}_W \right\}$$

is closed under addition and closed under scalar multiplication. (Note that $\ker(T)$ is non–empty because $\vec{0}_V \in \ker(T)$.)

Let $\vec{x}$ and $\vec{y}$ be vectors in $\ker(T)$. Then we know that $T(\vec{x}) = \vec{0}_W$ and $T(\vec{y}) = \vec{0}_W$. Since T is a linear transformation, we have

$$T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y}) = \vec{0}_W + \vec{0}_W = \vec{0}_W$$

and thus $\vec{x}+\vec{y} \in \ker(T)$. We have shown that $\ker(T)$ is closed under addition.

Let $\vec{x}$ be a vector in V and let c be a scalar. Then we know that $T(\vec{x}) = \vec{0}_W$. Since T is a linear transformation, then

$$T(c\vec{x}) = cT(\vec{x}) = (c)\vec{0}_W = \vec{0}_W$$

and thus $c\vec{x} \in \ker(T)$. We have shown that $\ker(T)$ is closed under scalar multiplication.

Exercise 5.6.8

Explanation: $D : C^1(R) \to C^0(R)$ is not invertible because for any function $g \in C^0(R)$, there are infinitely many functions $f \in C^1(R)$ in such that $D(f) = g$. For example,

$$D(5x+12) = 5$$

and

$$D(5x-37) = 5.$$

As another example

$$D(\sin(x)) = \cos(x)$$

and

$$D(\sin(x)+14) = \cos(x).$$

Exercise 5.6.10

Answer: Since

$$S(\langle a_1, a_2, a_3, \ldots\rangle) = \langle a_2, a_3, a_4, \ldots\rangle,$$

then

$$\begin{aligned} S^2(\langle a_1, a_2, a_3, \ldots\rangle) &= S(S(\langle a_1, a_2, a_3, \ldots\rangle)) \\ &= S(\langle a_2, a_3, a_4, \ldots\rangle) \\ &= \langle a_3, a_4, a_5, \ldots\rangle \end{aligned}$$

and

$$\begin{aligned} S^3(\langle a_1, a_2, a_3, \ldots\rangle) &= S\left(S^2(\langle a_1, a_2, a_3, \ldots\rangle)\right) \\ &= S(\langle a_3, a_4, a_5, \ldots\rangle) \\ &= \langle a_4, a_5, a_6, \ldots\rangle. \end{aligned}$$

Exercise 5.6.12

a) $[2+2x+2x^2]_{\mathcal{B}} = \langle 2,2,2\rangle$

c) $[x]_{\mathcal{B}} = \langle 0,1,0\rangle$

e) $[\langle 1,-2,-2\rangle]_{\mathcal{B}}^{-1} = 1-2x-2x^2$

Exercise 5.6.14

a)

$$[-3\sin(x)+2\cos(x)+3x\sin(x)-3x\cos(x)]_{\mathcal{B}} = \langle -3,2,3,-3\rangle$$

c)

$$[-4\cos(x)+5x\sin(x)+x\cos(x)+3\sin(x)]_{\mathcal{B}} = \langle 3,-4,5,1\rangle .$$

e) $[\,\langle 0,0,1,0\rangle]_{\mathcal{B}}^{-1} = x\sin(x)$

Exercise 5.6.15 Note that

$$\begin{aligned} D(1) &= 0 \\ D(x) &= 1 \\ D\left(x^2\right) &= 2x \\ D\left(x^3\right) &= 3x^2. \end{aligned}$$

Hence

$$\begin{aligned} [D(1)]_{\mathcal{B}} &= \langle 0,0,0,0\rangle \\ [D(x)]_{\mathcal{B}} &= \langle 1,0,0,0\rangle \\ \left[D\left(x^2\right)\right]_{\mathcal{B}} &= \langle 0,2,0,0\rangle \\ \left[D\left(x^3\right)\right]_{\mathcal{B}} &= \langle 0,0,3,0\rangle . \end{aligned}$$

The matrix of D with respect to the basis $\mathcal{B}$ is

$$A_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since

$$\text{rref}\,(A_{\mathcal{B}}) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

we see that $\text{Range}\,(D_{\mathcal{B}})$ has basis

$$\{\langle 1,0,0,0\rangle, \langle 0,2,0,0\rangle, \langle 0,0,3,0\rangle\}$$

and this translates, via the inverse coordinate transformation, to

$$\{1, 2x, 3x^2\}$$

being a basis for $\text{Range}\,(D)$. Hence

$$\text{Range}\,(D) = \text{Span}\,\{1, 2x, 3x^2\}$$

and $\dim(\text{Range}\,(D)) = 3$. Since $\text{Span}\,\{1, 2x, 3x^2\}$ is the same thing as $\text{Span}\,\{1, x, x^2\}$, we could also say that

$$\text{Range}\,(D) = \text{Span}\,\{1, x, x^2\} = P_2.$$

The row reduction we have done also shows that every vector in $\ker(D_{\mathcal{B}})$ must have the form $\langle t, 0, 0, 0\rangle$, where t can be any scalar, and this means that

$$\ker(D_{\mathcal{B}}) = \text{Span}\,\{\langle 1,0,0,0\rangle\}.$$

This translates, via the inverse coordinate transformation, to

$$\ker(D) = \text{Span}\,\{1\}.$$

We see that the Fundamental Theorem of Linear Algebra is satisfied:

$$\dim(\text{Range}\,(D)) + \dim(\ker(D)) = 3 + 1 = 4 = \dim(P_3).$$

Now observe that

$$A_{\mathcal{B}}^2 = \left(\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}\right)^2 = \begin{bmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_{\mathcal{B}}^3 = \left(\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right)^3 = \begin{bmatrix} 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_{\mathcal{B}}^3 = \left(\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right)^4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

which makes sense because for any function $f \in P_3$ we have $D^4(f) = z$.

Exercise 5.6.17

Solution: First we compute

$$\begin{aligned} D(1) &= 0 \\ D(e^x \sin(x)) &= e^x \sin(x) + e^x \cos(x) \\ D(e^x \cos(x)) &= -e^x \sin(x) + e^x \cos(x) \end{aligned}$$

from which we see that

$$\begin{aligned} [D(1)]_{\mathcal{B}} &= \langle 0, 0, 0 \rangle \\ [D(e^x \sin(x))]_{\mathcal{B}} &= \langle 0, 1, 1 \rangle \\ [D(e^x \cos(x))]_{\mathcal{B}} &= \langle 0, -1, 1 \rangle . \end{aligned}$$

The matrix of D with respect to $\mathcal{B}$ is thus

$$A_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix}.$$

This means that the matrix of the fifth derivative transformation, D^5, with respect to $\mathcal{B}$ is

$$A_{\mathcal{B}}^5 = \left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix} \right)^5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -4 & 4 \\ 0 & -4 & -4 \end{bmatrix}$$

The coordinate vector of the function $f(x) = 5 - e^x \sin(x)$ with respect to the ordered basis $\mathcal{B}$ is

$$[f]_{\mathcal{B}} = [5 - e^x \sin(x)]_{\mathcal{B}} = \langle 5, -1, 0 \rangle .$$

Since

$$A_{\mathcal{B}}^{5}[f]_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -4 & 4 \\ 0 & -4 & -4 \end{bmatrix} \langle 5, -1, 0 \rangle = \langle 0, 4, 4 \rangle$$

then

$$D^{5}(5 - e^{x}\sin(x)) = 4e^{x}\sin(x) + 4e^{x}\cos(x).$$

Exercise 5.6.20 **Solution:** The matrix we found in Exercise 5.6.17 is

$$A_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix}.$$

If we are asked to evaluate the indefinite integral

$$\int e^{x}\sin(x)\ dx,$$

then what we are being asked is to find all functions F such that $D(F) = e^{x}\sin(x)$. Since the coordinate vector of $f(x) = e^{x}\sin(x)$ with respect to the ordered basis

$$\mathcal{B} = \{1, e^{x}\sin(x), e^{x}\cos(x)\}$$

is

$$[f]_{\mathcal{B}} = [e^{x}\sin(x)]_{\mathcal{B}} = \langle 0, 1, 0 \rangle,$$

then we need to solve the equation $A_{\mathcal{B}}[F]_{\mathcal{B}} = [f]_{\mathcal{B}}$ for F. If we let the unknown $[F]_{\mathcal{B}} = \langle c_1, c_2, c_3 \rangle$, then the equation we want to solve is

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix} \langle c_1, c_2, c_3 \rangle = \langle 0, 1, 0 \rangle.$$

We form the augmented matrix for this equation and row reduce to obtain

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

This tells us that the solutions of $A_{\mathcal{B}}[F]_{\mathcal{B}} = [f]_{\mathcal{B}}$ are

$$[F]_{\mathcal{B}} = \left\langle t, \frac{1}{2}, -\frac{1}{2} \right\rangle$$

where t can be any scalar. Hence the solutions of $D(F) = e^x \sin(x)$ are

$$t(1) + \frac{1}{2}e^x \sin(x) - \frac{1}{2}e^x \cos(x)$$

where t can be any scalar. If we write C instead of t, we have

$$\int e^x \sin(x)\ dx = \frac{1}{2}e^x \sin(x) - \frac{1}{2}e^x \cos(x) + C.$$

Chapter 5 Additional Exercises

1. The **identity transformation** $E : R^2 \to R^2$ is defined by

$$E(\vec{x}) = \vec{x}.$$

   (a) Show that E satisfies both of the requirements of Definition 5.2.1 and is thus a linear transformation.

   Answer: Let $\vec{x}$ and $\vec{y}$ be any two vectors in R^2 and let c be any scalar. Then

$$E(\vec{x} + \vec{y}) = \vec{x} + \vec{y} = E(\vec{x}) + E(\vec{y})$$

   and

$$E(c\vec{x}) = c\vec{x} = cE(\vec{x}).$$

   This shows that E satisfies both requirements of Definition 5.2.1 and is thus a linear transformation.

   (b) Suppose that L is any line in R^2. To what line does the identity transformation map L? In other words, what is $E(L)$?

   Answer: Since E does not do anything to vectors in R^2, then $E(L) = L$ for any line L.

2. The **zero transformation** $Z : R^2 \to R^2$ is defined by

$$Z(\vec{x}) = \vec{0}_2.$$

   Let us show that the zero transformation defined by $Z(\vec{x}) = \vec{0}_2$ for $\vec{x} \in R^2$ is a linear transformation.

Let $\vec{x}$ and $\vec{y}$ be any two vectors in R^2 and let c be any scalar. Then

$$Z(\vec{x}+\vec{y}) = \vec{0}_2 = \vec{0}_2 + \vec{0}_2 = Z(\vec{x}) + Z(\vec{y})$$

and

$$Z(c\vec{x}) = \vec{0}_2 = c\vec{0}_2 = cZ(\vec{x}),$$

so Z is a linear transformation.

Z maps every point in R^2 to the zero vector. Thus if L is any line in R^2, then the entire line L gets mapped to $\vec{0}_2$. Basically what Z does is to "squash" all of R^2 onto a single point.

4. For the following linear transformations $T : R^n \to R^m$, determine the range of T and the kernel of T. Also determine whether or not T is invertible. If T is invertible, then find the formula for T^{-1}.

   a. $T : R^2 \to R^2$ defined by $T(\langle x_1, x_2\rangle) = \langle -4x_1 + 2x_2, -4x_1 - 6x_2\rangle$
   Solution: The matrix of T and its rref are

$$A = \begin{bmatrix} -4 & 2 \\ -4 & -6 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \text{rref}(A).$$

Since the pivot columns of A are a basis for $\text{Range}(T)$, then $\{\langle -4, -4\rangle, \langle 2, -6\rangle\}$ is a basis for $\text{Range}(T)$ and

$$\text{Range}(T) = \text{Span}\{\langle -4, -4\rangle, \langle 2, -6\rangle\} = R^2.$$

To find $\ker(T)$, we solve $A\vec{x} = \vec{0}_2$ using the augmented matrix

$$\left[\begin{array}{cc|c} -4 & 2 & 0 \\ -4 & -6 & 0 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right].$$

By the above row reduction, we see that the only solution of $A\vec{x} = \vec{0}_2$ is $\vec{x} = \vec{0}_2$. Thus

$$\ker(T) = \left\{\vec{0}_2\right\}.$$

Since T maps R^2 onto R^2 (because every row of A contains a pivot) and T is one–to–one (because every column of A contains a pivot), then T is invertible. The matrix for T^{-1} is

$$A^{-1} = \begin{bmatrix} -\frac{3}{16} & -\frac{1}{16} \\ \frac{1}{8} & -\frac{1}{8} \end{bmatrix}.$$

Thus $T^{-1}(\vec{x}) = A^{-1}\vec{x}$ for all $\vec{x}$ in R^2. The formula for T^{-1} is

$$T^{-1}(\langle x_1, x_2\rangle) = \left\langle -\frac{3}{16}x_1 - \frac{1}{16}x_2, \frac{1}{8}x_1 - \frac{1}{8}x_2 \right\rangle.$$

b. $T : R^2 \to R^2$ defined by $T(\langle x_1, x_2\rangle) = \langle -4x_1 + 2x_2, -4x_1 + 2x_2\rangle$
Solution: The matrix of T and its rref are

$$A = \begin{bmatrix} -4 & 2 \\ -4 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 0 \end{bmatrix} = \text{rref}(A).$$

Since the pivot columns of A are a basis for Range (T), then $\{\langle -4, -4\rangle\}$ is a basis for Range (T) and

$$\text{Range}(T) = \text{Span}\{\langle -4, -4\rangle\}.$$

To find $\ker(T)$, we solve $A\vec{x} = \vec{0}_2$ using the augmented matrix

$$\left[\begin{array}{cc|c} -4 & 2 & 0 \\ -4 & -6 & 0 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{array}\right].$$

By the above row reduction, we see that the solutions of $A\vec{x} = \vec{0}_2$ are

$$x_1 = \frac{1}{2}t$$
$$x_2 = t.$$

Thus the solutions are

$$\vec{x} = t\left\langle \frac{1}{2}, 1 \right\rangle$$

and we see that

$$\ker(T) = \text{Span}\left\{\left\langle \frac{1}{2}, 1 \right\rangle\right\}.$$

Since T does not map R^2 onto R^2 (because not every row of A contains a pivot) and T is not one–to–one (because not every column of A contains a pivot), then T is not invertible.

d. $T : R^5 \to R^2$ defined by $T(\langle x_1, x_2, x_3, x_4, x_5 \rangle) = \langle 5x_1 - 3x_2 - 3x_3 - 5x_4 - 2x_5, x_3 \rangle$

Solution: The matrix of T and its rref are

$$A = \begin{bmatrix} 5 & -3 & -3 & -5 & -2 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & -\frac{3}{5} & 0 & -1 & -\frac{2}{5} \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} = \text{rref}(A).$$

Since the pivot columns of A are a basis for Range(T), then

$$\{\langle 5, 0 \rangle, \langle -3, 1 \rangle\}$$

is a basis for Range(T) and

$$\text{Range}(T) = \text{Span}\{\langle 5, 0 \rangle, \langle -3, 1 \rangle\} = R^2.$$

To find ker(T), we solve $A\vec{x} = \vec{0}_2$ using the augmented matrix

$$\left[\begin{array}{ccccc|c} 5 & -3 & -3 & -5 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right] \to \left[\begin{array}{ccccc|c} 1 & -\frac{3}{5} & 0 & -1 & -\frac{2}{5} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right].$$

By the above row reduction, we see that the solutions of $A\vec{x} = \vec{0}_2$ are

$$\begin{aligned} x_1 &= \frac{3}{5}r + t + \frac{2}{5}s \\ x_2 &= r \\ x_3 &= 0 \\ x_4 &= t \\ x_5 &= s \end{aligned}$$

which we can write in the vector form

$$\vec{x} = r\left\langle \frac{3}{5}, 1, 0, 0, 0 \right\rangle + t\langle 1, 0, 0, 1, 0 \rangle + s\left\langle \frac{2}{5}, 0, 0, 0, 1 \right\rangle.$$

Thus

$$\ker(T) = \text{Span}\left\{\left\langle \frac{3}{5}, 1, 0, 0, 0 \right\rangle, \langle 1, 0, 0, 1, 0 \rangle, \left\langle \frac{2}{5}, 0, 0, 0, 1 \right\rangle\right\}.$$

T maps R^5 onto R^2 (because every row of A contains a pivot). T is not one–to–one (because not every column of A contains a pivot). Since T is not one–to–one, then T is not invertible.

5. Find the linear transformation that reflects vectors in R^2 through the x_1 axis. To do this

   (a) Determine $T(\langle 1,0\rangle)$ and $T(\langle 0,1\rangle)$.
   Answer:
$$\begin{aligned} T(\langle 1,0\rangle) &= \langle 1,0\rangle \\ T(\langle 0,1\rangle) &= \langle 0,-1\rangle \end{aligned}$$

   (b) Use what you found in part a to write down the matrix A such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in R^2$.
   Answer: The matrix for T is
$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

   (c) Write the formula for T in the form $T(\langle x_1, x_2\rangle) = \langle \_\_\_\_, \_\_\_\_\rangle$.
   Answer: The formula for T is $T(\langle x_1, x_2\rangle) = \langle x_1, -x_2\rangle$.

7. **Solution:** Referring to Figure 5.26, the transformation that rotates the line L by θ clockwise $R_{-\theta}$, which has matrix
$$A_{-\theta} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}.$$

   $R_{-\theta}$ rotates L onto the x_1 axis.

   The transformation, S, that reflects vectors through the x_1 axis has matrix
$$A_S = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

   The transformation, R_θ, that rotates the line x_1 axis by θ counterclockwise has matrix
$$A_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

   The desired reflection (reflecting vectors through the line L) is achieved by the composition $T = R_\theta \circ S \circ R_{-\theta}$ which has matrix
$$A_T = A_\theta S A_{-\theta} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}.$$

First note that

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

and thus

$$\begin{aligned} A_T &= \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \qquad (A.6) \\ &= \begin{bmatrix} \cos^2(\theta) - \sin^2(\theta) & 2\cos(\theta)\sin(\theta) \\ 2\cos(\theta)\sin(\theta) & \sin^2(\theta) - \cos^2(\theta) \end{bmatrix} \\ &= \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}. \end{aligned}$$

8. Use the general result that you found in Exercise 7 to find the linear transformation $T : R^2 \to R^2$ that reflects vectors through the following lines L:

   (a) L is the line that makes an angle of $60°$ with the positive x_1 axis.

   Solution: Using the matrix (A.6) from Exercise 7, we see that the transformation that reflects vectors through the line, L, that makes an angle of $60°$ with the positive x_1 axis is

$$A_T = \begin{bmatrix} \cos(2(60°)) & \sin(2(60°)) \\ \sin(2(60°)) & -\cos(2(60°)) \end{bmatrix} = \begin{bmatrix} -1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}.$$

   (b) L is the line $x_2 = 2x_1$.

   Solution: The line $L : x_2 = 2x_1$ is pictured in Figure A.8.

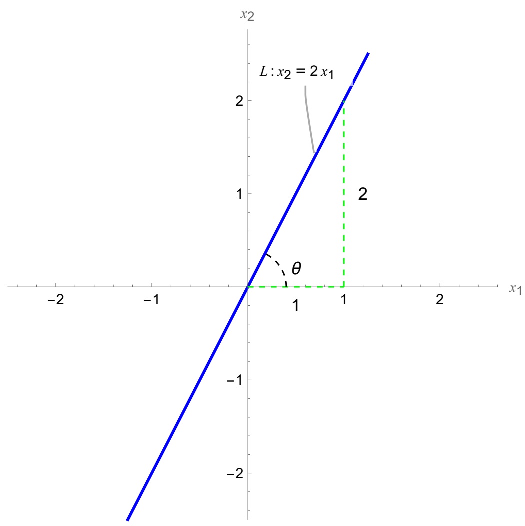


Figure A.8: $L : x_2 = 2x_1$

The angle that this line makes with the positive x_1 axis is θ where

$$\tan(\theta) = \frac{2}{1}.$$

Thus

$$\theta = \arctan(2).$$

We also see from the picture that

$$\begin{aligned}\sin(\theta) &= \frac{2}{\sqrt{5}} \\ \cos(\theta) &= \frac{1}{\sqrt{5}}.\end{aligned}$$

This gives

$$\begin{aligned}\sin(2\theta) &= 2\sin(\theta)\cos(\theta) \\ &= 2\left(\frac{2}{\sqrt{5}}\right)\left(\frac{1}{\sqrt{5}}\right) \\ &= \frac{4}{5}\end{aligned}$$

and

$$\begin{aligned}\cos(2\theta) &= \cos^2(\theta) - \sin^2(\theta) \\ &= \left(\frac{1}{\sqrt{5}}\right)^2 - \left(\frac{2}{\sqrt{5}}\right)^2 \\ &= -\frac{3}{5}.\end{aligned}$$

The matrix for T is thus

$$A_T = \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix} = \begin{bmatrix} -\frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix}.$$

The formula for T is

$$T(\langle x_1, x_2\rangle) = \left\langle -\frac{3}{5}x_1 + \frac{4}{5}x_2, \frac{4}{5}x_1 + \frac{3}{5}x_2 \right\rangle.$$

As a way to check that this formula is correct, what if we take a point $(x_1, 2x_1)$ that is actually on the line L? For this point, the formula gives us

$$\begin{aligned}T(\langle x_1, x_2\rangle) &= T(\langle x_1, 2x_1\rangle) \\ &= \left\langle -\frac{3}{5}x_1 + \frac{4}{5}(2x_1), \frac{4}{5}x_1 + \frac{3}{5}(2x_1)\right\rangle \\ &= \left\langle -\frac{3}{5}x_1 + \frac{8}{5}x_1, \frac{4}{5}x_1 + \frac{6}{5}x_1\right\rangle \\ &= \langle x_1, 2x_1\rangle \\ &= \langle x_1, x_2\rangle\end{aligned}$$

which makes sense because a point that is already on L stays where it is when T acts on it.

10. **Solution:** We want to show that the matrices

$$A = \begin{bmatrix} 3 & 0 \\ 1 & 3 \end{bmatrix} \text{ and } B = \begin{bmatrix} 7/2 & -1/2 \\ 1/2 & 5/2 \end{bmatrix}$$

are similar to each other. To do this we need to find an by finding an invertible 2×2 matrix C such that $CA = BC$. To do this, we will let

$$C = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$$

be an unknown matrix and solve the equation $CA = BC$. This equation is

$$\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 7/2 & -1/2 \\ 1/2 & 5/2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}.$$

Matrix multiplication on the left had side of the above equation gives

$$\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3x_1 + x_2 & 3x_2 \\ 3x_3 + x_4 & 3x_4 \end{bmatrix}$$

and matrix multiplication on the right hand side of the above equation gives

$$\begin{bmatrix} 7/2 & -1/2 \\ 1/2 & 5/2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} = \begin{bmatrix} \frac{7}{2}x_1 - \frac{1}{2}x_3 & \frac{7}{2}x_2 - \frac{1}{2}x_4 \\ \frac{1}{2}x_1 + \frac{5}{2}x_3 & \frac{1}{2}x_2 + \frac{5}{2}x_4 \end{bmatrix}.$$

Thus we need to solve

$$\begin{bmatrix} 3x_1 + x_2 & 3x_2 \\ 3x_3 + x_4 & 3x_4 \end{bmatrix} = \begin{bmatrix} \frac{7}{2}x_1 - \frac{1}{2}x_3 & \frac{7}{2}x_2 - \frac{1}{2}x_4 \\ \frac{1}{2}x_1 + \frac{5}{2}x_3 & \frac{1}{2}x_2 + \frac{5}{2}x_4 \end{bmatrix}.$$

We have a system of four equations with four unknowns:

$$\begin{aligned} 3x_1 + x_2 &= \frac{7}{2}x_1 - \frac{1}{2}x_3 \\ 3x_2 &= \frac{7}{2}x_2 - \frac{1}{2}x_4 \\ 3x_3 + x_4 &= \frac{1}{2}x_1 + \frac{5}{2}x_3 \\ 3x_4 &= \frac{1}{2}x_2 + \frac{5}{2}x_4 \end{aligned}$$

which we can write as the homogeneous system

$$\begin{array}{rcrcrcrcl} -\frac{1}{2}x_1 & + & x_2 & + & \frac{1}{2}x_3 & & & = & 0 \\ & & -\frac{1}{2}x_2 & & & + & \frac{1}{2}x_4 & = & 0 \\ -\frac{1}{2}x_1 & & & + & \frac{1}{2}x_3 & + & x_4 & = & 0 \\ & & -\frac{1}{2}x_2 & & & + & \frac{1}{2}x_4 & = & 0 \end{array}.$$

The augmented matrix for this system and its rref are

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

We see that the system has infinitely many solutions which are given by

$$\begin{aligned} x_1 &= t + 2s \\ x_2 &= s \\ x_3 &= t \\ x_4 &= s. \end{aligned}$$

Thus the matrix we are looking for has the form

$$C = \begin{bmatrix} t+2s & s \\ t & s \end{bmatrix}.$$

We need to be sure to choose t and s so that C is invertible. Let us choose $t = 0$ and s=1. Then we get

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

and we see that C is invertible with

$$C^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 \end{bmatrix}.$$

We now check that $CA = BC$:

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

and

$$BC = \begin{bmatrix} 7/2 & -1/2 \\ 1/2 & 5/2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 3 \\ 1 & 3 \end{bmatrix},$$

so it works! We have shown that the matrices A and B are similar to each other.

11. **Answer:** In order to answer this question, we need to use some facts from calculus. These facts are:

    (a) If the sequence $\vec{a}$ converges and has limit $L_{\vec{a}}$ and the sequence $\vec{b}$ converges and has limit $L_{\vec{b}}$, then the sequence $\vec{a}+\vec{b}$ also converges and has limit $L_{\vec{a}} + L_{\vec{b}}$.

(b) If the sequence $\vec{a}$ converges and has limit $L_{\vec{a}}$ and c is any scalar, then the sequence $c\vec{a}$ converges and has limit $cL_{\vec{a}}$

The above two facts are what is needed to conclude that C is a subspace of R^{∞}.

Now we will show that $T : C \to R$ defined by $T(\vec{a}) = L_{\vec{a}}$ is a linear transformation.

Let $\vec{a}$ and $\vec{b}$ be in C. Then

$$T\left(\vec{a}+\vec{b}\right) = L_{\vec{a}+\vec{b}} = L_{\vec{a}} + L_{\vec{b}} = T(\vec{a}) + T\left(\vec{b}\right),$$

which shows that the first requirement of Definition 5.6.1 is satisfied.

Let $\vec{a}$ be in C and let c be a scalar. then

$$T(c\vec{a}) = L_{c\vec{a}} = cL_{\vec{a}} = cT(\vec{a}),$$

which shows that the second requirement of Definition 5.6.1 is satisfied. Therefore T is a linear transformation.

If we are given any real number, r, then we can easily come up with an infinite sequence whose limit is r. In fact that sequence $\vec{a} = \langle r, r, r, \ldots \rangle$ has limit r. This shows that $\text{Range}(T) = R$ and that T maps C onto R.

$\ker(T)$ is the set of all $\vec{a} \in C$ for which $T(\vec{a}) = 0$. In other words, $\ker(T)$ is the set of all convergent infinite sequences that have 0 as their limit. It is certainly not true that T is one–to–one, because given any real number r, we can find infinitely many different sequences whose limit is r. For example, if

$$\vec{a} = \langle 0, 0, 0, \ldots \rangle \text{ (all components are 0)}$$

and

$$\vec{b} = \left\langle 1, \frac{1}{2}, \frac{1}{3}, \ldots \right\rangle \text{ (nth component is $1/n$)},$$

then $T(\vec{a}) = 0$ and $T\left(\vec{b}\right) = 0$. Clearly T is not invertible.

13. **Solution:** Since

$$D(\sin(x)) = \cos(x)$$
$$D(\cos(x)) = -\sin(x),$$

then

$$[D(\sin(x))] = \langle 0, 1\rangle$$
$$[D(\cos(x))] = \langle -1, 0\rangle .$$

This tells us that the matrix of D with respect to the ordered basis $\mathcal{B} = \{\sin(x), \cos(x)\}$ is

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

We obtain

$$A^2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$
$$A^3 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$
$$A^4 = I_2.$$

The coordinate vector of $f(x) = \sin(x)$ with respect to $\mathcal{B}$ is $\langle 1, 0\rangle$. Since

$$A\langle 1,0\rangle = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\langle 1,0\rangle = \langle 0,1\rangle$$
$$A^2\langle 1,0\rangle = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}\langle 1,0\rangle = \langle -1,0\rangle$$
$$A^3\langle 1,0\rangle = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\langle 1,0\rangle = \langle 0,-1\rangle$$
$$A^4\langle 1,0\rangle = I_2\langle 1,0\rangle = \langle 1,0\rangle ,$$

then

$$D(\sin(x)) = [\langle 0,1\rangle]^{-1} = \cos(x)$$
$$D^2(\sin(x)) = [\langle -1,0\rangle]^{-1} = -\sin(x)$$
$$D^3(\sin(x)) = [\langle 0,-1\rangle]^{-1} = -\cos(x)$$
$$D^4(\sin(x)) = [\langle 1,0\rangle]^{-1} = \sin(x).$$

A.6 Chapter 6 Exercises:

Exercise 6.0.1 For the matrix $A = \begin{bmatrix} 5 & -1 \\ 3 & 1 \end{bmatrix}$,

1. Evaluate $A\vec{x}$ where $\vec{x} = \langle 1, 1 \rangle$.
   Answer:
   $$A\vec{x} = \langle \text{Row}_1(A) \cdot \vec{x}, \text{Row}_2(A) \cdot \vec{x} \rangle = \langle 4, 4 \rangle.$$

2. Show that if $\vec{x}$ is any vector in $\text{Span}\{\langle 1, 1 \rangle\}$, then $A\vec{x} = 4\vec{x}$.
   Answer: We saw above that $A\langle 1, 1 \rangle = \langle 4, 4 \rangle = 4\langle 1, 1 \rangle$. If $\vec{x}$ is any vector in $\text{Span}\{\langle 1, 1 \rangle\}$, then $\vec{x} = c\langle 1, 1 \rangle$ for some scalar c. By properties of the matrix-vector product,
   $$A\vec{x} = A(c\langle 1, 1 \rangle) = cA\langle 1, 1 \rangle = c(4\langle 1, 1 \rangle) = 4(c\langle 1, 1 \rangle) = 4\vec{x}.$$

3. Identify the matrix $A - 4I_2$, and show that this matrix is not invertible.
   Answer:
   $$A - 4I_2 = \begin{bmatrix} 5 & -1 \\ 3 & 1 \end{bmatrix} - \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 3 & -3 \end{bmatrix}$$
   Note that $\text{rref}(A - 4I_2) = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \neq I_2$ which shows that $A - 4I_2$ is not invertible.

Exercise 6.0.2 Consider the matrix $A = \begin{bmatrix} 4 & 7 \\ 2 & -1 \end{bmatrix}$.

1. 1 Find a nonzero vector $\vec{v} = \langle v_1, v_2 \rangle$ such that $A\vec{v} = 6\vec{v}$.
   Answer: This can be set up as a system of equations.
   $$A\vec{v} = \langle 4v_1 + 7v_2, 2v_1 - v_2 \rangle = \langle 6v_1, 6v_2 \rangle, \qquad \begin{array}{rcrcl} 4v_1 & + & 7v_2 & = & 6v_1 \\ 2v_1 & - & v_2 & = & 6v_2 \end{array}.$$
   Moving the v_1 and v_2 to the left side, we get a conventional looking system that happens to be homogeneous.
   $$\begin{array}{rcrcl} -2v_1 & + & 7v_2 & = & 0 \\ 2v_1 & - & 7v_2 & = & 0 \end{array}.$$
   This can be solved using a matrix with row reduction (note that the matrix is not the same as A). Solutions will be $\vec{v} = t\left\langle \frac{7}{2}, 1 \right\rangle$, for any real t. Taking any nonzero value for t will give a correct solution.

2. Confirm that $A\vec{x} = 6\vec{x}$ for every vector in $\text{Span}\{\vec{v}\}$, where $\vec{v}$ is the vector you found in part 1. above.
   Answer: We can use any example for $\vec{v}$ found above. A nice choice is to take $t = 2$ to avoid fractions. This gives $\vec{v} = \langle 7, 2\rangle$. We can do the product $A\vec{v}$.

$$A\vec{v} = \langle 4(7) + 7(2), 2(7) + (-1)(2)\rangle = \langle 42, 12\rangle = 6\langle 7, 2\rangle = 6\vec{v}.$$

   Now, let $\vec{x} = c\vec{v}$. Then

$$A\vec{x} = A(c\vec{v}) = cA\vec{v} = c(6\vec{v}) = 6(c\vec{v}) = 6\vec{x}.$$

3. 3 Compute the matrix $A - (-3)I_2$.
   Answer:

$$A - (-3)I_2 = \begin{bmatrix} 4 & 7 \\ 2 & -1 \end{bmatrix} - \begin{bmatrix} -3 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 7 & 7 \\ 2 & 2 \end{bmatrix}.$$

4. 4 Find a basis for $\mathcal{N}(A-(-3)I_2)$, i.e., the null space of the matrix that you computed in part 3. above.
   Answer: We can row reduce $\left[A - (-3)I_2 \,|\, \vec{0}_2\right]$.

$$\left[\begin{array}{cc|c} 7 & 7 & 0 \\ 2 & 2 & 0 \end{array}\right] \quad \xrightarrow{rref} \quad \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

   So if $(A - (-3)I_2)\vec{x} = \vec{0}_2$, then $x_1 = -x_2$ with x_2 free. Such a vector

$$\vec{x} = t\langle -1, 1\rangle, \quad t \in R.$$

   So we can take as a basis $\{\langle -1, 1\rangle\}$.

5. Show that if $\vec{x}$ is in $\mathcal{N}(A - (-3)I_2)$, then $A\vec{x} = -3\vec{x}$. (Hint: start by taking $\vec{x}$ to be the basis element you found in part 4. above.)
   Answer: Taking the hint, let's note that

$$A\langle -1, 1\rangle = \langle 4(-1) + 7(1), 2(-1) + (-1)(1)\rangle = \langle 3, -3\rangle = -3\langle -1, 1\rangle.$$

   So if $\vec{x}$ is any vector in $\mathcal{N}(A - (-3)I_2) = \text{Span}\,\{\langle -1, 1\rangle\}$, then $\vec{x} = c\langle -1, 1\rangle$ for some scalar c, and

$$A\vec{x} = Ac\langle -1, 1\rangle = cA\langle -1, 1\rangle = c\,(-3\langle -1, 1\rangle) = -3(c\langle -1, 1\rangle) = -3\vec{x}.$$

Exercise 6.0.3 Diagonal matrices are particularly easy to work with. Consider the 3×3 diagonal matrix $A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$ with a, b, and c some real numbers. Show that there are three vectors, say $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_3$, such that

$$A\vec{v}_1 = a\vec{v}_1, \quad A\vec{v}_2 = b\vec{v}_2, \quad \text{and} \quad A\vec{v}_3 = c\vec{v}_3.$$

Answer: We might just recall that $A\vec{e}_i = \mathrm{Col}_i(A)$, and use $\vec{v}_1 = \vec{e}_1$, $\vec{v}_2 = \vec{e}_2$ and $\vec{v}_3 = \vec{e}_3$. It is easy to confirm that

$$A\vec{e}_1 = a\vec{e}_1, \quad A\vec{e}_2 = b\vec{e}_2, \quad \text{and} \quad A\vec{e}_3 = c\vec{e}_3.$$

Exercise 6.1.1 Evaluate the determinant of each of the matrices

1. $A = \begin{bmatrix} 2 & -4 \\ 6 & 10 \end{bmatrix}$
   Answer: $\det(A) = 2(10) - (-4)(6) = 44$

2. $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$
   Answer: $\det(B) = 0(0) - (-1)(1) = 1$

3. $C = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ where θ is a real number.
   Answer: $\det(C) = \cos\theta(\cos\theta) - (-\sin\theta)\sin\theta = \cos^2\theta + \sin^2\theta = 1$

Exercise 6.1.2

1. Show that for $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $\det(2A) = 4\det(A)$.
   Answer: $\det(2A) = \det\left(\begin{bmatrix} 2a & 2b \\ 2c & 2d \end{bmatrix}\right) = 2a(2d) - 2b(2c) = 4ad - 4bc = 4(ad - bc) = 4\det(A)$.

2. Show that for $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $\det(3A) = 9\det(A)$.
   Answer: $\det(3A) = \det\left(\begin{bmatrix} 3a & 3b \\ 3c & 3d \end{bmatrix}\right) = 3a(3d) - 3b(3c) = 9ad - 9bc = 9(ad - bc) = 9\det(A)$.

3. Can you make a conjecture about the relationship between $\det(kA)$ and $\det(A)$ for a 2×2 matrix A and a scalar k?
Answer: The value of $\det(kA)$ should be k^2 times the value $\det(A)$. This is clear given that

$$\det(kA) = ka(kd) - kb(kc) = k^2(ad - bc) = k^2 \det(A).$$

Exercise 6.1.3 Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, and suppose $\det(A) \neq 0$. Show that

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Answer: We can form the product and show that the result is I_2. Note that

$$\begin{aligned} \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \frac{1}{\det(A)} \begin{bmatrix} ad - bc & bd - bd \\ -ca + ac & -cb + ad \end{bmatrix} \\ &= \frac{1}{\det(A)} \begin{bmatrix} \det(A) & 0 \\ 0 & \det(A) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

Exercise 6.1.4 Evaluate the determinant of each 3×3 matrix.

1. $A = \begin{bmatrix} 1 & 2 & -1 \\ 4 & 3 & 0 \\ -2 & 1 & 5 \end{bmatrix}$
Answer: $\det(A) = -35$

2. $A = \begin{bmatrix} -3 & 4 & 3 \\ 3 & -4 & -3 \\ 2 & 1 & 0 \end{bmatrix}$
Answer: $\det(A) = 0$

3. $A = \begin{bmatrix} -5 & -1 & 1 \\ 2 & 1 & 1 \\ 3 & 1 & -1 \end{bmatrix}$
Answer: $\det(A) = 4$

Exercise 6.1.5 Suppose $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$. Show that $\det(A) = a_{11}a_{22}a_{33}$.

Answer: We can use Definition 6.1.2. Note that

$$A_{11} = \begin{bmatrix} a_{22} & a_{23} \\ 0 & a_{33} \end{bmatrix}, \quad \text{so} \quad \det(A_{11}) = a_{22}a_{33} - 0 = a_{22}a_{33}.$$

$$A_{12} = \begin{bmatrix} 0 & a_{23} \\ 0 & a_{33} \end{bmatrix}, \quad \text{so} \quad \det(A_{12}) = 0(a_{33}) - 0(a_{23}) = 0.$$

$$A_{13} = \begin{bmatrix} 0 & a_{22} \\ 0 & 0 \end{bmatrix}, \quad \text{so} \quad \det(A_{13}) = 0(0) - 0(a_{22}) = 0.$$

So.

$$\begin{aligned} \det(A) &= a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13}) \\ &= a_{11}(a_{22}a_{33}) - a_{12}(0) + a_{13}(0) \\ &= a_{11}a_{22}a_{33} \end{aligned}$$

Exercise 6.1.6 Find the determinant of the matrix $A = \begin{bmatrix} 1 & 2 & -1 \\ 3 & 4 & 0 \\ 2 & -2 & 3 \end{bmatrix}$ by computing a cofactor expansion

1. across the second row, **Answer:** $\det(A) = 8$
2. down the first column, **Answer:** $\det(A) = 8$
3. across the third row. **Answer:** $\det(A) = 8$

Exercise 6.1.7 Find the determinant of each matrix using a cofactor expansion that minimizes the computations.

1. $A = \begin{bmatrix} 1 & 0 & -1 & 2 \\ 2 & 0 & 4 & -3 \\ 0 & 2 & 5 & 2 \\ 1 & 0 & -1 & 0 \end{bmatrix}$ **Answer:** The second column only has one nonzero entry, so cofactor expansion down this second column requires the least amout of computation. $\det(A) = 24$

2. $B = \begin{bmatrix} 3 & -4 & 0 \\ 0 & -6 & 0 \\ 1 & 1 & 0 \end{bmatrix}$ **Answer:** If the third column is used for the cofactor expansion, it isn't necessary to compute any cofactors. $\det(B) = 0$

Exercise 6.1.8 Suppose A is an $n \times n$ matrix, and A has a row or a column vector of all zeros. Explain why $\det(A) = 0$.
Answer: If we choose the row or column of zeros for the cofactor expansion, then each factor a_{ij} will be zero.

Exercise 6.1.9 Confirm each of the three statements in Property 6.4 for a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.
Answer: Here is an example of a row replacement. The other two properties can be examined in a similar manner. Let's obtain a new matrix B by performing the operation $kR_1 + R_2 \to R_2$. Then

$$B = \begin{bmatrix} a & b \\ ka + c & kb + d \end{bmatrix}$$

making

$$\det(B) = a(kb + d) - (ka + c)b = akb + ad - kab - bc = ad - bc = \det(A).$$

Exercise 6.1.10 Suppose A is a 4×4 matrix that is row equivalent to the matrix

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

If the following row operations were performed on A to produce B, determine $\det(A)$.

- $-2R_1 + R_2 \to R_2$
- $R_3 \leftrightarrow R_4$
- $3R_2 + R_3 \to R_3$

- $\frac{1}{2}R_3 \to R_3$
- $-R_2 + R_4 \to R_4$

Answer: There are three row replacements that do not change the determinant. The row swap ($R_3 \leftrightarrow R_4$) will introduce a factor of -1 and the scaling ($\frac{1}{2}R_3 \to R_3$) will introduce a factor of $\frac{1}{2}$. So

$$\det(B) = (-1)\left(\frac{1}{2}\right)\det(A)., \quad \text{hence} \quad \det(A) = -2\det(B).$$

Since B is triangular, its determinant is just the product of the diagonal entries. So

$$\det(A) = -2\det(B) = -2(3)(4)(-1)(-2) = -48.$$

Exercise 6.1.11 If A is an $n \times n$ matrix, explain why $\det(kA) = k^n \det(A)$ for scalar k.
Answer: The matrix kA is obtained by multiplying every row by k. With n rows, this is equivalent to performing the scaling $kR_i \to R_i$ n times, and each one gives a factor of k. So

$$\det(kA) = \underbrace{k \cdot k \cdots k}_{n \text{ factors}} \det(A) = k^n \det(A).$$

Exercise 6.1.12 For each pair of matrices A and B, evaluate the products AB and BA. Compute the determinants $\det(A)$, $\det(B)$, $\det(AB)$, and $\det(BA)$ and confirm that $\det(AB) = \det(A)\det(B) = \det(BA)$.

1. $A = \begin{bmatrix} 1 & 2 \\ -4 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 1 \\ 2 & 5 \end{bmatrix}$.
   Answer: Note that $AB = \begin{bmatrix} 7 & 11 \\ -6 & 11 \end{bmatrix}$ and $BA = \begin{bmatrix} -1 & 9 \\ -18 & 19 \end{bmatrix}$.

$$\det(A) = 1(3) - (-4)(2) = 11, \quad \det(B) = 3(5) - 2(1) = 13,$$

$$\det(AB) = 7(11) - (-6)(11) = 143,$$

$$\text{and} \quad \det(BA) = -1(-19) - (-18)(9) = 143.$$

Sure enough, $\det(A)\det(B) = 11(13) = 143 = \det(AB) = \det(BA)$.

2. $A = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & 1 \\ 2 & -2 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 6 & 0 & -1 \\ 2 & 2 & 2 \\ -1 & 0 & 3 \end{bmatrix}$. **Answer:**

$$AB = \begin{bmatrix} 2 & -2 & 3 \\ 5 & 6 & 9 \\ 4 & -4 & 6 \end{bmatrix}, \quad \text{and} \quad BA = \begin{bmatrix} 4 & -4 & 8 \\ 6 & 0 & 14 \\ 5 & -5 & 10 \end{bmatrix}.$$

$$\det(A) = 0, \quad \det(B) = 34, \quad \det(AB) = 0, \quad \det(BA) = 0.$$

Again, $\det(A)\det(B) = (0)(34) = 0 = \det(AB) = \det(BA)$.

Exercise 6.1.13 For each matrix A, determine all values of λ, if any, such that A is not invertible.

1. $A = \begin{bmatrix} 2-\lambda & 1 \\ 5 & -2-\lambda \end{bmatrix}$
   Answer: The determinant $\det(A) = (2-\lambda)(-2-\lambda) - 1(5) = \lambda^2 - 9$. A is not invertible if $\det(A) = 0$. This gives two values of λ,

$$\lambda^2 - 9 = 0 \quad \text{if} \quad \lambda = 3, \quad \text{or} \quad \lambda = -3.$$

2. $A = \begin{bmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{bmatrix}$
   Answer: $\det(A) = (1-\lambda)^2 - 1(1) = \lambda^2 - 2\lambda$. So $\det(A) = 0$ if $\lambda = 0$ or $\lambda = 2$.

3. $A = \begin{bmatrix} 3-\lambda & 0 \\ 2 & 3-\lambda \end{bmatrix}$
   Answer: $\det(A) = (3-\lambda)^2 - 2(0) = (3-\lambda)^2$. The only value of λ such that $\det(A) = 0$ is $\lambda = 3$.

4. $A = \begin{bmatrix} 2-\lambda & 4 \\ -1 & 3-\lambda \end{bmatrix}$
   Answer: $\det(A) = (2-\lambda)(3-\lambda) - (-1)(4) = \lambda^2 - 5\lambda + 10$. The descriminant of this quadratic is $(-5)^2 - 4(1)(10) = -15$. So there are no real numbers λ for which $\det(A) = 0$. So A is invertible for all real λ. (Depending on the context, we may be interested in complex solutions. There are two complex roots of the quadratic, $\lambda = \frac{5 \pm i\sqrt{15}}{2}$. Substituting these numbers into the matrix A would result in complex entries.)

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

   Answer: Taking a cofactor expansion down the first column,

   $$\det(A) = (1-\lambda)((3-\lambda)^2-(-1)(-1)) = (1-\lambda)(\lambda^2-6\lambda+8) = (1-\lambda)(\lambda-2)(\lambda-4).$$

   There are three values of λ for which $\det(A) = 0$, i.e., for which A is not invertible. $\lambda = 1$, $\lambda = 2$ or $\lambda = 4$. (**Tip:** Since our goal is to solve an equation of the form "*some expression in* $\lambda = 0$", it's to our advantage to factor rather than multiply everything out.)

Exercise 6.2.1 Let $A = \begin{bmatrix} 3 & -1 \\ 5 & -3 \end{bmatrix}$.

1. Show that $\lambda = 2$ is an eigenvalue of A by finding a nonzero vector $\vec{x}$ such that $A\vec{x} = 2\vec{x}$.
   Answer: Following the procedure seen in Example 6.2.1, we find that $\vec{x} = t\langle 1, 1\rangle$. Any choice of $t \neq 0$ will give a valid solution.

2. Show that $\vec{x} = \langle 1, 5\rangle$ is an eigenvector of A by finding a scalar λ such that $A\vec{x} = \lambda\vec{x}$.
   Answer: We find that $A\vec{x} = \langle 3(1) - 1(5), 5(1) - 3(5)\rangle = \langle -2, -10\rangle = -2\langle 1, 5\rangle$. So $\langle 1, 5\rangle$ is an eigenvector corresponding to the eigenvalue $\lambda = -2$.

3. Show that the number $\lambda = 3$ is not an eigenvalue of A. (Hint: Show that $A\vec{x} = 3\vec{x}$ has no nontrivial solutions.)
   Answer: If we let $\vec{x} = \langle x_1, x_2\rangle$ and set up $A\vec{x} = 3\vec{x}$, we can rearrange to get a homogeneous system.

   $$\begin{array}{rcrcl} 3x_1 & - & x_2 & = & 3x_1 \\ 5x_1 & - & 3x_2 & = & 3x_2 \end{array} \quad \text{becomes} \quad \begin{array}{rcrcl} & - & x_2 & = & 0 \\ 5x_1 & - & 6x_2 & = & 0 \end{array}.$$

   Note that performing row reduction on the corresponding augmented matrix gives

   $$\left[\begin{array}{cc|c} 0 & -1 & 0 \\ 5 & -6 & 0 \end{array}\right] \xrightarrow{rref} \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right].$$

   So the only solution is $\vec{x} = \langle 0, 0\rangle$ which can not be an eigenvector. (**Tip:** If we mistake a number for an eigenvalue, this sort of result will alert us to the error.)

Exercise 6.2.2 We've seen that if $(\lambda, \vec{x})$ is an eigenvalue-eigenvector pair for a matrix A, then $A\vec{x}$ is in Span$\{\vec{x}\}$. Consider the transformation $R_{90^\circ}(\vec{x}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \vec{x}$ that rotates a vector in R^2 by 90° counterclockwise. Explain why there are no (real) numbers λ that are eigenvalue of the matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

Answer: We recall that R_{90° is invertible (by the clockwise rotation R_{-90°), so there are no nontrivial solutions to the equation $R_{90^\circ}(\vec{x}) = \vec{0}_2$. That is, there's no eigenvector associated with the number $\lambda = 0$, so zero is not an eigenvalue. For any nonzero λ the nonzero vector $\vec{x}$ and $\lambda\vec{x}$ are parallel. But $\vec{x}$ and $R_{90^\circ}(\vec{x})$ are perpendicular since the rotation is 90°. So it's not possible for $R_{90^\circ}(\vec{x})$ to be in Span$\{\vec{x}\}$.

Exercise 6.2.3 For each matrix, determine all eigenvalues and for each eigenvalue, find a corresponding eigenvector.

1. $A = \begin{bmatrix} 2 & 1 \\ 5 & -2 \end{bmatrix}$
   Answer: The characteristic polynomial is $P_A(\lambda) = \lambda^2 - 9$ with roots $\lambda_1 = 3$ and $\lambda_2 = -3$. Solving the homogeneous equation $(A - 3I_2)\vec{x} = \vec{0}_2$ gives solutions $\vec{x} = t\langle 1, 1\rangle, t \in R$. Solving the homogeneous equation $(A - (-3)I_2)\vec{x} = \vec{0}_2$ gives solutions $\vec{x} = s\left\langle -\frac{1}{5}, 1\right\rangle, s \in R$. Example eigenvectors can be chosen by selecting any nonzero value of the parameter (t or s). Selecting $t = 1$ and $s = 5$, we have eigenvalue-eigenvector pairs,
   $$\lambda_1 = 3, \quad \vec{x}_1 = \langle 1, 1\rangle, \quad \text{and} \quad \lambda_2 = -3, \quad \vec{x}_2 = \langle -1, 5\rangle.$$

2. $A = \begin{bmatrix} 3 & 0 \\ 2 & 3 \end{bmatrix}$
   Answer: The characteristic polynomial is $P_A(\lambda) = (3 - \lambda)^2$. There is one root, hence one eigenvalue $\lambda = 3$. The solutions of the homogeneous equation $(A - 3I_2)\vec{x} = \vec{0}_2$ are of the form $\vec{x} = t\langle 0, 1\rangle$, $t \in R$. We get the one eigenvalue-eigenvector pair
   $$\lambda = 3, \quad \vec{x} = \langle 0, 1\rangle.$$

3. $A = \begin{bmatrix} 1 & 2 & -2 \\ 0 & 3 & -1 \\ 0 & -1 & 3 \end{bmatrix}$
Answer: The characteristic polynomial is $P_A(\lambda) = (1-\lambda)(\lambda-2)(\lambda-4)$. We find three eigenvalues $\lambda_1 = 1$, $\lambda_2 = 2$ and $\lambda_3 = 4$. The equation $(A - 1I_3)\vec{x} = \vec{0}_3$ has solutions $\vec{x} = t\langle 1, 0, 0\rangle$, $t \in R$. The equation $(A - 2I_3)\vec{x} = \vec{0}_3$ has solutions $\vec{x} = s\langle 0, 1, 1\rangle$, $s \in R$. And the equation $(A - 4I_3)\vec{x} = \vec{0}_3$ has solutions $\vec{x} = u\left\langle -\frac{4}{3}, -1, 1\right\rangle$, $u \in R$. Selecting $t = s = 1$ and $u = 3$, we have three eigenvalue-eigenvector pairs.

$$\lambda_1 = 1, \quad \vec{x}_1 = \langle 1, 0, 0\rangle, \quad \lambda_2 = 2, \quad \vec{x}_2 = \langle 0, 1, 1\rangle,$$

$$\text{and} \quad \lambda_3 = 4, \quad \vec{x}_3 = \langle -4, -3, 3\rangle.$$

Exercise 6.2.4 Consider the pair of matrices

$$A = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 5 \end{bmatrix}.$$

1. Find the characteristic polynomials P_A and P_B and show that they are equal, $P_A(\lambda) = P_B(\lambda)$.
Answer: Fortunately, these matrices are triangular, so the determinants will be easy to take.

$$\det(A - \lambda I_3) = \det\left(\begin{bmatrix} 3-\lambda & 1 & 0 \\ 0 & 3-\lambda & 0 \\ 0 & 0 & 5-\lambda \end{bmatrix}\right) = (3-\lambda)^2(5-\lambda)$$

and

$$\det(B - \lambda I_3) = \det\left(\begin{bmatrix} 3-\lambda & 0 & 0 \\ 0 & 3-\lambda & 1 \\ 0 & 0 & 5-\lambda \end{bmatrix}\right) = (3-\lambda)^2(5-\lambda).$$

The common polynomial can be expanded to

$$P_A(\lambda) = P_B(\lambda) = -\lambda^3 + 11\lambda^2 - 39\lambda + 45,$$

but it isn't necessary to do so.

2. Identify the eigenvalues of A and for each eigenvalue of A determine its algebraic multiplicity and its geometric multiplicity.
   Answer: From the factored form of P_A, there are two eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 5$. Finding bases for the eigenspaces, we find

$$\text{for } \lambda_1 = 3, \text{ a basis for the eigenspace is } \{\langle 1,0,0\rangle\},$$

$$\text{for } \lambda_2 = 5, \text{ a basis for the eigenspace is } \{\langle 0,0,1\rangle\}.$$

   For $\lambda_1 = 3$, the algebraic multiplicity is two and the geometric multiplicity is one. For $\lambda_2 = 5$, the algebraic multiplicity is one and the geometric multiplicity is one.

3. Identify the eigenvalues of B and for each eigenvalue of B determine its algebraic multiplicity and its geometric multiplicity.
   Answer: From the factored form of P_B, B has the same two eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 5$. Finding bases for the eigenspaces, we find

$$\text{for } \lambda_1 = 3, \text{ a basis for the eigenspace is } \{\langle 1,0,0\rangle, \langle 0,1,0\rangle\},$$

$$\text{for } \lambda_2 = 5, \text{ a basis for the eigenspace is } \{\langle 0,1,2\rangle\}.$$

   For $\lambda_1 = 3$, the algebraic multiplicity is two and the geometric multiplicity is also two. For $\lambda_2 = 5$, the algebraic multiplicity is one and the geometric multiplicity is one.

Exercise 6.2.5 For each of the matrices

$$A = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 5 \end{bmatrix}.$$

from Exercise 6.2.4, construct an eigenbasis or explain why one does not exist.

Answer: We recall that both matrices had the same two eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 5$ where λ_1 has algebraic multiplicity 2 and λ_2 has algebraic multiplicity 1. For A, the geometric multiplicity of both eigenvalues is 1. A does not give rise to an eigenbasis because the sum of the geometric multiplicities is $1 + 1 = 2 \neq 3$.

For B, we found that the geometric multiplicity of $\lambda_1 = 3$ is 2, and we found a basis for the eigenspace $E_B(3)$ to be $\{\langle 1,0,0\rangle, \langle 0,1,0\rangle\}$. For the eigenvalue $\lambda_2 = 5$, we found a basis for the corresponding eigenspace $E_B(3)$ to be $\{\langle 0,1,2\rangle\}$. The sum of the geometric multiplicities is $2+1=3$, and we have three linearly independent eigenvectors. An eigenbasis for the matrix B is

$$\mathcal{E}_B = \{\langle 1,0,0\rangle, \langle 0,1,0\rangle, \langle 0,1,2\rangle\}.$$

Exercise 6.3.1 Let $B = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 5 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}$. Show that B is diagonalizable. To do this, find C^{-1} and compute the product $C^{-1}BC$.
Answer: C^{-1} can be computed by row reducing $\left[C \,|\, I_3\right]$ (it can even be done by hand with only a couple of row operations). The inverse

$$C^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} \end{bmatrix}.$$

Then

$$\begin{aligned} C^{-1}BC &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 5 \\ 0 & 0 & 10 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix} \end{aligned}$$

We see that the product $C^{-1}BC$ is a diagonal matrix $D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$ which shows that B is diagonalizable.

Exercise 6.3.2 For each matrix, either diagonalize the matrix (i.e., identify the diagonal matrix D and invertible matrix C) or show that the matrix is not diagonalizable.

1. $A = \begin{bmatrix} -4 & 7 \\ -2 & 5 \end{bmatrix}$
   Answer: Note that answers may vary in the order in which the eigenvalues appear in D. The order of the columns in C and values in D should be consistent. A possible answer is
$$D = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1 & 7 \\ 1 & 2 \end{bmatrix}$$

2. $L = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$
   Answer: This matrix is not diagonalizable. The only eigenvalue is $\lambda = 2$ with algebraic multiplicity 2 and geometric mutliplicity 1.

3. $H = \begin{bmatrix} 2 & 3 \\ -1 & 5 \end{bmatrix}$
   Answer: The matrix does not have any real eigenvalues. The characteristic equation is $\lambda^2 - 7\lambda + 13 = 0$, which does not have real roots.

4. $B = \begin{bmatrix} -2 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 6 \end{bmatrix}$
   Answer: This matrix is not diagonalizable. The eigenvalues are $\lambda_1 = -2$ and $\lambda_2 = 6$. The algebraic multiplicity of λ_1 is two, but its geometric mutliplicity is one.

5. $G = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 0 & 0 & -1 \end{bmatrix}$
   Answer: Answers may vary. An answer is
$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 2 \end{bmatrix}$$

Exercise 6.3.3

1. Find a 3×3 matrix A having eigenvalues $L = \{1, -4, 5\}$ and for which $\mathcal{E}_A = \{\langle 1, 1, 3\rangle, \langle 1, 1, -3\rangle, \langle 0, -1, -2\rangle\}$ is an eigenbasis.

Answer: Since we have an eigenbasis, our matrix A would be diagonalizable. We can create the diagonal matrix from the known eigenvalues and the invertible matrix from the known eigenvectors. Let

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

A solution can be computed as $A = CDC^{-1}$.

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & -1 \\ 3 & -3 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & -1 \\ 3 & -3 & -2 \end{bmatrix}^{-1} = \frac{1}{6} \begin{bmatrix} 1 & -10 & 5 \\ -29 & 20 & 5 \\ -33 & 78 & -9 \end{bmatrix}.$$

2. Is your answer A in part 1. above unique? That is, can you find another 3×3 matrix having eigenvalues $L = \{1, -4, 5\}$ and eigenbasis $\mathcal{E}_A = \{\langle 1, 1, 3\rangle, \langle 1, 1, -3\rangle, \langle 0, -1, -2\rangle\}$?
Hint: Use technology to try rearranging the columns of D and C. Do different arrangements give you the same matrix A?

Exercise 6.4.1 Consider the ordered basis of R^2 given by $\mathcal{C} = \{\langle 1, 1\rangle, \langle -1, 5\rangle\}$.

1. Identify the change of basis matrix C and its inverse C^{-1}.
Answer: The matrix and its inverse are

$$C = \begin{bmatrix} 1 & -1 \\ 1 & 5 \end{bmatrix}, \quad \text{and} \quad C^{-1} = \frac{1}{6} \begin{bmatrix} 5 & 1 \\ -1 & 1 \end{bmatrix}.$$

(Because C is a 2×2 matrix, C^{-1} can be found by doing row reduction on $[C \mid I_2]$ or by using the handy formula from Exercise 6.1.3.)

2. Find the coordinate vectors relative to the basis $\mathcal{C}$ for the following vectors.
Answer: To get the coordinate vectors, we use the equation $[\vec{x}]_{\mathcal{C}} = C^{-1}\vec{x}$ (some might be done by observation without any computations).

   (a) $\vec{x} = \langle 1, 1\rangle \quad [\vec{x}]_{\mathcal{C}} = \langle 1, 0\rangle$

   (b) $\vec{y} = \langle -1, 5\rangle \quad [\vec{y}]_{\mathcal{C}} = \langle 0, 1\rangle$

   (c) $\vec{z} = \langle 0, 1\rangle \quad [\vec{z}]_{\mathcal{C}} = \left\langle \frac{1}{6}, \frac{1}{6}\right\rangle$

3. Find the representation relative to the standard basis for the vectors having the given coordinate vectors relative to the basis $\mathcal{C}$.
   Answer: To get the coordinate vectors, we use the equation $\vec{x} = C[\vec{x}]_{\mathcal{C}}$

   (a) $[\vec{u}]_{\mathcal{C}} = \langle 1, 1\rangle \quad \vec{u} = \langle 0, 6\rangle$
   (b) $[\vec{v}]_{\mathcal{C}} = \langle -1, 5\rangle \quad \vec{v} = \langle -6, 9\rangle$
   (c) $[\vec{w}]_{\mathcal{C}} = \langle 0, 1\rangle \quad \vec{w} = \langle -1, 5\rangle$

Exercise 6.4.2 Let $T : R^2 \to R^2$ be the linear transformation defined by $T(\vec{x}) = A\vec{x}$ where $A = \begin{bmatrix} -2 & 6 \\ -2 & 5 \end{bmatrix}$. Find a basis $\mathcal{C}$ of R^2 such that the $\mathcal{C}$-matrix of T is diagonal. Find the $\mathcal{C}$-matrix.
Answer: The basis $\mathcal{C}$ will have to be an eigenbasis for A. The characteristic polynomial $P_A(\lambda) = \det(A - \lambda I_2) = \lambda^2 - 3\lambda + 2$ which has two roots, $\lambda_1 = 1$ and $\lambda_2 = 2$ with corresponding eigenvectors $\vec{x}_1 = \langle 2, 1\rangle$ and $\vec{x}_2 = \langle 3, 2\rangle$. So a basis is $\mathcal{C} = \{\langle 2, 1\rangle, \langle 3, 2\rangle\}$. The change of basis matrix for this would be $C = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$. The inverse is $C^{-1} = \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix}$. The $\mathcal{C}$-matrix for T would be

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

Note: Answers can vary in some details. For example, the order of the eigenvalues in the diagonal $\mathcal{C}$-matrix can be swapped. The eigenvectors chosen may be different, but the eigenvector for each eigenvalue should be some scalar multiple of the ones used here.

Exercise 6.4.3 A matrix A is called **symmetric** if $A = A^T$. It is known that symmetric matrices are always diagonalizable. Moreover, the eigenvectors for distinct eigenvalues are orthogonal. That is, a symmetric matrix has an eigenbasis of mutually orthogonal vectors. Let $T : R^3 \to R^3$ be the linear transformation defined by $T(\vec{x}) = A\vec{x}$ for the matrix A given below. Find a basis $\mathcal{C}$ of R^3 such that the $\mathcal{C}$-matrix of T is diagonal, and confirm that the basis elements are orthogonal. Find the $\mathcal{C}$-matrix.

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

Answer: The matrix A does satisfy $A^T = A$. The determinant of $A-\lambda I_3$ can be taken across the third row to simplify the calculation.

$$P_A(\lambda) = (-2-\lambda)((6-\lambda)^2-1) = (-2-\lambda)(\lambda^2-12\lambda+35) = (-2-\lambda)(\lambda-5)(\lambda-7).$$

We can find an eigenvector for each of the three eigenvalues, $\lambda_1 = -2$, $\lambda_2 = 5$ and $\lambda_3 = 5$. Finding a basis for the null space of $A - \lambda_i I_3$ in the usual way, we can find corresponding eigenvectors $\vec{x}_1 = \langle 0,0,1\rangle$, $\vec{x}_2 = \langle -1,1,0\rangle$ and $\vec{x}_3 = \langle 1,1,0\rangle$. So a basis is

$$\mathcal{C} = \{\langle 0,0,1\rangle, \langle -1,1,0\rangle, \langle 1,1,0\rangle\}.$$

To show that the basis vectors are orthogonal, we can check the dot products,

$$\begin{aligned}\vec{x}_1 \cdot \vec{x}_2 &= 0(-1) + 0(1) + 1(0) = 0,\\ \vec{x}_1 \cdot \vec{x}_3 &= 0(1) + 0(1) + 1(0) = 0,\\ \vec{x}_2 \cdot \vec{x}_3 &= -1(1) + 1(1) + 0(0) = 0.\end{aligned}$$

The change of basis matrix and its inverse are

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

The diagonal $\mathcal{C}$-matrix of T is

$$C^{-1}AC = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 7 \end{bmatrix}.$$

Note: Answers may vary in some details. The eigenvectors should be orthogonal despite variations.

Exercise 6.4.4 Let $T : R^2 \to R^2$ be the shear transformation such that $T(\vec{e}_1) = \vec{e}_1 - 2\vec{e}_2$ and $T(\vec{e}_2) = \vec{e}_2$ (so T leaves $\vec{e}_2$ fixed). Determine whether there is a basis $\mathcal{C}$ of R^2 such that the $\mathcal{C}$-matrix of T is diagonal. If so, find the diagonal matrix.
Answer: From the description,

$$T(\vec{e}_1) = \langle 1,0\rangle - 2\langle 0,1\rangle = \langle 1,-2\rangle, \quad \text{and} \quad T(\vec{e}_2) = \langle 0,1\rangle.$$

So the standard matrix

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

We can already see the eigenvalues of A since it is lower triangular. There's one eigenvalue $\lambda = 1$. It must have algebraic multiplicity two. We can confirm that by noting that $P_A(\lambda) = (1-\lambda)^2$. We also already know that one eigenvector is $\vec{e}_2$ because $T(\vec{e}_2) = 1\vec{e}_2$. A will be diagonalizable if there is a second, linearly independent eigenvector. If we look for a basis for $\mathcal{N}(A - 1I_2)$, we find

$$\left[A - 1I_2 \,|\, \vec{0}_2\right] \quad \xrightarrow{rref} \quad \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

So if $\vec{x} = \langle x_1, x_2 \rangle$ is a solution of $(A - 1I_2)\vec{x} = \vec{0}_2$, then $x_1 = 0$ and x_2 is free. This only gives the one basis vector $\vec{e}_2$. So there is no basis for R^2 for which the matrix for T is diagonal.

Chapter 6 Additional Exercises

1. If $A = [a_{11}]$ is a 1×1 matrix, we define its determinant to be $\det(A) = a_{11}$. Use this definition to show that the determinant of a 2×2 matrix from Definition 6.1.1 is the same as a cofactor expansion

$$\det(A) = \sum_{j=1}^{2} (-1)^{1+j} a_{1j} \det(A_{1j}).$$

Answer: If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then notice that $A_{11} = [d]$, and $A_{12} = [c]$. So

$$\begin{aligned} \det(A) &= (-1)^{1+1} a \det(A_{11}) + (-1)^{1+2} b \det(A_{12}) \\ &= 1(a)(d) + (-1)(b)(c) \\ &= ad - bc. \end{aligned}$$

2. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Suppose A has two (not necessarily distinct) eigenvalues λ_1 and λ_2. Show that

$$a + d = \lambda_1 + \lambda_2 \quad \text{and} \quad \det(A) = \lambda_1 \lambda_2.$$

(Hint: The characteristic polynomial must factor as $P_A(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)$. Compare this to P_A obtained in the usual way.)
Answer: On the one hand, the characteristic polynomial is

$$P_A(\lambda) = \det\left(\begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix}\right) = (a-\lambda)(d-\lambda)-bc = \lambda^2-(a+d)\lambda+ad-bc.$$

On the other hand, we have the characteristic polynomial from its roots

$$P_A(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2.$$

Comparing the two representations for the same polynomial, we can equate the coefficient of λ to get

$$-(a+d) = -(\lambda_1 + \lambda_2), \quad \text{i.e.,} \quad a + d = \lambda_1 + \lambda_2.$$

And equating the constant terms, we get

$$ad - bc = \lambda_1\lambda_2, \quad \text{i.e.,} \quad \det(A) = \lambda_1\lambda_2.$$

3. Give a coherent argument that if $A = [a_{ij}]$ is an $n \times n$ triangular matrix, the eigenvalues of A are its diagonal entries, a_{ii}.
Answer: The critical observation here is that if A is triangular, then $A - \lambda I_n$ is also triangular, since this matrix difference only has the effect of subtracting λ from each diagonal entry of A—it has no effect on the entries off of the main diagonal. So when we take $\det(A - \lambda I_n)$, we can use Property 6.3 that says that the determinant will be the product of the diagonal entries. This gives

$$P_A(\lambda) = (a_{11} - \lambda)(a_{22} - \lambda)\cdots(a_{nn} - \lambda).$$

From this factored form, we see that the zeros of P_A, which are the eigenvalues of the matrix A, are the numbers a_{ii} from the main diagonal.

4. Suppose A is an invertible matrix. Show that $\det(A^{-1}) = (\det(A))^{-1}$. That is, show that the determinant of A^{-1} is the reciprocal of the determinant of A.
Answer: We recall that for $n \times n$ matrices A and B, $\det(AB) = \det(A)\det(B)$. We also note that the $n \times n$ identity matrix is a diagonal

matrix with all diagonal entries equal to 1. So $\det(I_n) = 1^n = 1$. Now, note that

$$AA^{-1} = I_n \quad \text{so that} \quad \det(AA^{-1}) = \det(I_n) = 1.$$

But we have $\det(AA^{-1}) = \det(A)\det(A^{-1})$. Since A is invertible, we know that $\det(A) \neq 0$, and we can divide by $\det(A)$. We have

$$\det(A)\det(A^{-1}) = 1 \quad \text{whence} \quad \det(A^{-1}) = \frac{1}{\det(A)}.$$

5. For the matrix A, evaluate $\det(A)$. Find all of the eigenvalues of A and show that $\det(A)$ is equal to the product of the eigenvalues of A.

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

Answer: First, taking a cofactor expansion down the first column of A (to take advantage of the zeros)

$$\det(A) = -2(4(-1) - (-1)(6)) = -2(-4+6) = -4.$$

To find the eigenvalues, we set up $\det(A - \lambda I_3)$. This determinant can also be taking down the first column.

$$\begin{aligned} \det\left(\begin{bmatrix} -2-\lambda & 1 & 3 \\ 0 & 4-\lambda & -1 \\ 0 & 6 & -1-\lambda \end{bmatrix}\right) &= (-2-\lambda)\big[(4-\lambda)(-1-\lambda) - 6(-1)\big] \\ &= (-2-\lambda)\big[\lambda^2 - 3\lambda - 4 + 6\big] \\ &= (-2-\lambda)\big[\lambda^2 - 3\lambda + 2\big] \\ &= (-2-\lambda)(\lambda-1)(\lambda-2) \end{aligned}$$

From this factored form, we see that there are three eigenvalues $\lambda_1 = -2$, $\lambda_2 = 1$, and $\lambda_3 = 2$. As expected,

$$\lambda_1\lambda_2\lambda_3 = -2(1)(2) = -4 = \det(A).$$

6. Suppose the $n \times n$ matrix A has n not necessarily distinct real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Show that $\det(A) = \lambda_1\lambda_2\cdots\lambda_n$, that is, the determinant of A is the product of its eigenvalues.

(Hint: The characteristic polynomial can be written as a product of linear factors $P_A(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda)$. How are $\det(A)$ and $P_A(0)$ related?)
Answer: Since $P_A(\lambda) = \det(A - \lambda I_n)$, we have that

$$P_A(0) = \det(A - 0I_n) = \det(A).$$

Taking the hint about P_A written as a product of linear factors,

$$\det(A) = P_A(0) = (\lambda_1 - 0)(\lambda_2 - 0)\cdots(\lambda_n - 0) = \lambda_1\lambda_2\cdots\lambda_n.$$

Side note: It turns out, this is the case even if some of the eigenvalues are not real. In that case, the real entries in A ensure that complex eigenvalues will occur in conjugate pairs, so the product of all of them will be a real number and will be the determinant of A.

7. Suppose A is an $n \times n$ invertible matrix and λ_0 is an eigenvalue of A. Show that $\dfrac{1}{\lambda_0}$ is an eigenvalue of A^{-1}.
Answer: Since λ_0 is an eigenvalue, there is a nonzero vector $\vec{x}$ such that $A\vec{x} = \lambda_0\vec{x}$. If we multiply both sides by A^{-1}, we get

$$A^{-1}A\vec{x} = A^{-1}(\lambda_0\vec{x}) \quad \Longrightarrow \quad I_n\vec{x} = \lambda_0 A^{-1}\vec{x}.$$

Since A is invertible, λ_0 is nonzero, so we can divide both sides by λ_0 to get

$$\frac{1}{\lambda_0}\vec{x} = A^{-1}\vec{x}, \quad \text{i.e.,} \quad A^{-1}\vec{x} = \frac{1}{\lambda_0}\vec{x}.$$

This shows that $\dfrac{1}{\lambda_0}$ is an eigenvalue of A^{-1}. It even shows that the eigenspace for A associated with λ_0 is the same as the eigenspace for A^{-1} associated with $\dfrac{1}{\lambda_0}$

8. Suppose A is a 5×5 matrix with characteristic polynomial

$$P_A(\lambda) = (2 - \lambda)^2(4 - \lambda)(-1 - \lambda)(6 - \lambda).$$

For each question, either provide a short answer or explain why it is not possible to answer.

(a) Is A invertible? **Answer:** Yes. The eigenvalues are 2, 4, -1 and 6. Zero is not an eigenvalue of A so it is invertible.

(b) Evaluate $\det(A-2I_5)$ **Answer:** Since 2 is an eigenvalue, $\det(A-2I_5)=0$.

(c) Is A diagonalizable? **Answer:** We can't know without more information. The algebraic multiplicity of 2 is two. A will be diagonalizable if the geometric multiplicity is two, but that can't be determined from the characteristic polynomial alone.

(d) Is there a nonzero vector $\vec{x}$ in R^5 such that $A\vec{x}=-\vec{x}$? **Answer:** Yes. Since -1 is an eigenvalue, there is nonzero eigenvector $\vec{x}$ such that $A\vec{x}=-1\vec{x}=-\vec{x}$.

(e) What is $\det(A)$? **Answer:** From the last exercise, $\det(A)=P_A(0)=2^2(4)(-1)(6)=-96$.

(f) Is $\det(A-5I_5)=0$? **Answer:** No. If this was zero, then 5 would be an eigenvalue, but $5-\lambda$ is not a factor of P_A.

(g) Is there an eigenbasis of R^5 for A? **Answer:** We can't know. This is the same question about diagonalizability. We would need to know the geometric multiplicity of the eigenvalue 2, but there's not enough information.

9. Suppose A is a 5×5 matrix with characteristic polynomial

$$P_A(\lambda)=-\lambda(1-\lambda)(-1-\lambda)(2-\lambda)(7-\lambda).$$

For each question, either provide a short answer or explain why it is not possible to answer.

(a) Is A invertible? **Answer:** No. One of the factors is $0-\lambda$, so zero is an eigenvalue of A making it not invertible. From the factored polynomial, the eigenvalues are $0,1,-1,2$ and 7.

(b) Is $A-I_5$ invertible? **Answer:** No, it's not. Since 1 is an eigenvalue, $\det(A-1I_5)=0$. So $A-I_5$ is not invertible.

(c) Is A diagonalizable? **Answer:** Yes it is. There are five distinct eigenvalues, so A is guaranteed to be diagonalizable.

(d) Is there a nonzero vector $\vec{x}$ in R^5 such that $A\vec{x}=\vec{x}$? **Answer:** Yes. 1 is an eigenvalue so there must be a nonzero vector such that $A\vec{x}=1\vec{x}$.

(e) What is $\det(A)$? **Answer:** Since zero is an eigenvalue, $\det(A) = 0$.

(f) Is $A - 5I_5$ invertible? **Answer:** Yes, it is. Since 5 is not an eigenvalue, it must be that $\det(A - 5I_5) \neq 0$ making $A - 5I_5$ invertible.

(g) Is there an eigenbasis of R^5 for A? **Answer:** Yes. This is the same as the question about diagonalizability above. Five distinct eigenvalues for a 5×5 matrix ensures that there are five linearly independent eigenvectors to make a basis for R^5.

10. Find a 3×3 matrix A having eigenvalues $L = \{2, -1, 3\}$ and eigenbasis $\mathcal{E}_A = \{\langle 1, 0, 1\rangle, \langle -2, 1, 0\rangle, \langle 3, 1, 2\rangle\}$.
Answer: A solution can be formed as $A = CDC^{-1}$ with the columns of the matrix C set as the eigenvectors and the entries on the diagonal matrix D the eigenvalues. Taking them in the order given in the problem statement, one solution is

$$A = \begin{bmatrix} 1 & -2 & 3 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & -2 & 3 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix}^{-1} = \frac{1}{3}\begin{bmatrix} 3 & 12 & 3 \\ 4 & 5 & -4 \\ 2 & 4 & 4 \end{bmatrix}.$$

11. Prove Theorem 6.3.5. That is, show that if A and B are similar, then for positive integer n, A^n and B^n are also similar with the same similarity transformation matrix. (Hint: using induction.)
Answer: Suppose A and B are similar, so there exists and invertible matrix C such that $B = C^{-1}AC$. This is the base case (the case $n = 2$ is also shown before the statement of Theorem 6.3.5). Suppose that for some integer $k \geq 1$ that $B^k = C^{-1}A^kC$, then note that

$$\begin{aligned} B^{k+1} = BB^k &= (C^{-1}AC)(C^{-1}A^kC) \\ &= C^{-1}A(CC^{-1})A^kC \\ &= C^{-1}AI_nA^kC \\ &= C^{-1}AA^kC \\ &= C^{-1}A^{k+1}C, \end{aligned}$$

so B^{k+1} is similar to A^{k+1}. It follows that B^n is similar to A^n for all integers $n \geq 1$.

12. Suppose A and B are similar, invertible matrices. Show that A^{-1} and B^{-1} are similar and that A^T and B^T are similar.
Answer: Since A and B are similar, there is an invertible matrix C such that $B = C^{-1}AC$. We recall that the inverse of a product is the product of the inverses in the reverse order (i.e., $(XY)^{-1} = Y^{-1}X^{-1}$), and a matrix is the inverse of its inverse (i.e., $(C^{-1})^{-1} = C$). Take the inverse of both sides of the equation between A and B to get
$$B^{-1} = \left(C^{-1}AC\right)^{-1} = C^{-1}A^{-1}(C^{-1})^{-1} = C^{-1}A^{-1}C.$$
So B^{-1} is similar to A^{-1} with the same similarity transform matrix as A and B. We also recall that the transpose of a product is the product of the transposes in the reverse order (i.e., $(XY)^T = Y^TX^T$), and the inverse of the transpose is the transpose of the inverse (i.e., $(C^T)^{-1} = (C^{-1})^T$). Take the transpose of the equation between A and B to get
$$B^T = \left(C^{-1}AC\right)^T = C^TA^T(C^{-1})^T = C^TA^T(C^T)^{-1}.$$
So A^T and B^T are similar. If the matrix for the similarity transformation between A and B is C, then the matrix for the transformation between A^T and B^T is $(C^T)^{-1}$.

13. (Involves calculus) An interesting use of diagonalization arises in the solution of linear systems of differential equations. We know, for example, that the simple differential equation $\frac{dy}{dt} = ay$, with a a constant, has family of solutions $y(t) = e^{at}y_0$ where y_0 is a scalar (it is the value of $y(t)$ when $t = 0$). We can formulate a vector version of this simple equation with $\vec{y}(t) = \langle x(t), y(t)\rangle$, a vector valued function of t. The derivative is taken entry-wise, $\frac{d\vec{y}}{dt} = \left\langle \frac{dx}{dt}, \frac{dy}{dt}\right\rangle$. If A is a 2×2 matrix, we can consider the vector differential equation
$$\frac{d\vec{y}}{dt} = A\vec{y},$$
and propose a solution analogous to the scalar version, $\vec{y}(t) = e^{At}\vec{y}_0$. This requires giving meaning to an exponential e^{At} when A is a matrix. We can turn to a series representation. Recall that the exponential e^x can be expressed in terms of the series
$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

This suggests a way to give meaning to a matrix exponential. We can define

$$e^{At} = I_n + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \cdots = \sum_{n=0}^{\infty} \frac{t^n}{n!}A^n.$$

If D is a diagonal matrix, $D = \begin{bmatrix} d_{11} & 0 \\ 0 & d_{22} \end{bmatrix}$, then we can get a nice form for the exponential of the matrix,

$$e^{Dt} = \begin{bmatrix} e^{d_{11}t} & 0 \\ 0 & e^{d_{22}t} \end{bmatrix}.$$

Glossing over some technical issues, we can show that if $D = C^{-1}AC$, then $e^{At} = Ce^{Dt}C^{-1}$. Determine the matrix exponential e^{At} if $A = \begin{bmatrix} -2 & 6 \\ -2 & 5 \end{bmatrix}$. (Note this is the matrix from Exercise 6.4.2.) Show that $e^{A(0)} = I_2$, that is, when $t = 0$, the matrix exponential is the identity (this is analogous to the fact that $e^0 = 1$).
Answer: From the previous exercise with this matrix A, $D = C^{-1}AC$ where

$$D = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}, \quad \text{and} \quad C^{-1} = \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix}.$$

So

$$\begin{aligned} e^{At} &= \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} e^t & 0 \\ 0 & e^{2t} \end{bmatrix} \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2e^t & -3e^t \\ -e^{2t} & 2e^{2t} \end{bmatrix} \\ &= \begin{bmatrix} 4e^t - 3e^{2t} & -6e^t + 6e^{2t} \\ 2e^t - 2e^{2t} & -3e^t + 4e^{2t} \end{bmatrix} \end{aligned}$$

Note that when $t = 0$,

$$e^{A(0)} = \begin{bmatrix} 4e^0 - 3e^{2(0)} & -6e^0 + 6e^{2(0)} \\ 2e^0 - 2e^{2(0)} & -3e^0 + 4e^{2(0)} \end{bmatrix} = \begin{bmatrix} 4-3 & -6+6 \\ 2-2 & -3+4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2.$$

A.7 Chapter 7 Exercises:

Exercise 7.1.1 Let $S = \{\langle 1, -1, 3\rangle, \langle -1, 2, 1\rangle, \langle 7, 4, -1\rangle\}$. Show that S is an orthgonal set.

Answer: Calling the vectors $\vec{v}_1, \vec{v}_2$, and $\vec{v}_3$, in the order they appear, we already know from Example 7.1.2 that $\vec{v}_1 \cdot \vec{v}_2 = 0$. We need to show that $\vec{v}_1 \cdot \vec{v}_3 = 0$ and $\vec{v}_2 \cdot \vec{v}_3 = 0$ as well. Note that

$$\begin{aligned} \vec{v}_1 \cdot \vec{v}_3 &= 1(7) + (-1)(4) + 3(-1) = 0 \\ \vec{v}_2 \cdot \vec{v}_3 &= -1(7) + 2(4) + 1(-1) = 0 \end{aligned}$$

Hence we can conclude that S is an orthogonal set.

Exercise 7.1.2 Let $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ where

$$\vec{v}_1 = \langle 3, 0, -3, 1\rangle, \quad \vec{v}_2 = \langle 2, 1, 1, -3\rangle, \quad \text{and} \quad \vec{v}_3 = \langle 1, 5, 2, 3\rangle.$$

In Example 7.1.1, we determined that S is an orthogonal basis for Span(S). Use the formula for the weights from Theorem 7.1.2 to express $\vec{x} = \langle 3, 3, -2, 4\rangle$ as a linear combination of the elements of S, and confirm that your solution is correct.

Answer: To use the formula fromTheorem 7.1.2 for the weights, we need the dot product of $\vec{x}$ with each basis element as well as the dot product of each basis element with itself.

$$\begin{aligned} \vec{x} \cdot \vec{v}_1 &= 3(3) + 3(0) + (-2)(-3) + 4(1) = 19, \\ \vec{x} \cdot \vec{v}_2 &= 3(2) + 3(1) + (-2)(1) + 4(-3) = -5, \\ \vec{x} \cdot \vec{v}_3 &= 3(1) + 3(5) + (-2)(2) + 4(3) = 26, \end{aligned}$$

and

$$\begin{aligned} \vec{v}_1 \cdot \vec{v}_1 &= 3^2 + 0^2 + (-3)^2 + 1^2 = 19, \\ \vec{v}_2 \cdot \vec{v}_2 &= 2^2 + 1^2 + 1^2 + (-3)^2 = 15, \\ \vec{v}_3 \cdot \vec{v}_3 &= 1^2 + 5^2 + 2^2 + 3^2 = 39. \end{aligned}$$

The weights are

$$c_1 = \frac{19}{19} = 1, \quad c_2 = -\frac{5}{15} = -\frac{1}{3}, \quad \text{and} \quad c_3 = \frac{26}{39} = \frac{2}{3}.$$

Hence

$$\vec{x} = \vec{v}_1 - \frac{1}{3}\vec{v}_2 + \frac{2}{3}\vec{v}_3.$$

It's not necessarily obvious that $\vec{x}$ is in Span(S), but we can confirm by actually computing this linear combination.

$$\langle 3, 0, -3, 1\rangle - \frac{1}{3}\langle 2, 1, 1, -3\rangle + \frac{2}{3}\langle 1, 5, 2, 3\rangle =$$
$$\left\langle 3 - \frac{2}{3} + \frac{2}{3}, 0 - \frac{1}{3} + \frac{10}{3}, -3 - \frac{1}{3} + \frac{4}{3}, 1 - \left(-\frac{3}{3}\right) + \frac{6}{3}\right\rangle =$$
$$\langle 3, 3, -2, 4\rangle = \vec{x}.$$

Exercise 7.1.3 Show that the set $\{\langle 2, 2, 1\rangle, \langle -2, 1, 2\rangle, \langle 1, -2, 2\rangle\}$ is an orthogonal basis for R^3, and find an associated orthonormal basis by normalizing the vectors.
Answer: Let's call the vectors $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_3$ in the order written. To show that the set is an orthogonal basis, we need to confirm that they are orthogonal.

$$\begin{aligned} \vec{v}_1 \cdot \vec{v}_2 &= 2(-2) + 2(1) + 1(2) = 0, \\ \vec{v}_1 \cdot \vec{v}_3 &= 2(1) + 2(-2) + 1(2) = 0, \\ \vec{v}_2 \cdot \vec{v}_3 &= -2(1) + 1(-2) + 2(2) = 0. \end{aligned}$$

The set is orthogonal. Since it's orthogonal, it is linearly independent. And since it is a linearly independent set of three vectors in R^3, it is a basis for R^3. To obtain an orthonormal basis, we need to find the magnitudes.

$$\|\vec{v}_1\| = \sqrt{2^2 + 2^2 + 1^2} = 3, \quad \|\vec{v}_2\| = \sqrt{(-2)^2 + 1^2 + 2^2} = 3,$$
$$\text{and} \quad \|\vec{v}_3\| = \sqrt{1^2 + (-2)^2 + 2^2} = 3.$$

They all have the same magnitude. An orthonormal basis is

$$\left\{\left\langle \frac{2}{3}, \frac{2}{3}, \frac{1}{3}\right\rangle, \left\langle -\frac{2}{3}, \frac{1}{3}, \frac{2}{3}\right\rangle, \left\langle \frac{1}{3}, -\frac{2}{3}, \frac{2}{3}\right\rangle\right\}$$

Exercise 7.2.1 Consider the parallel vectors $\vec{v}_1 = \langle -2, 3\rangle$ and $\vec{v}_2 = \langle 4, -6\rangle$, and let $\vec{x} = \langle -5, 2\rangle$. Show that

$$\operatorname{proj}_{\vec{v}_1} \vec{x} = \operatorname{proj}_{\vec{v}_2} \vec{x}.$$

Answer:

$$\text{proj}_{\vec{v}_1}\vec{x} = \frac{\vec{x}\cdot\vec{v}_1}{\|\vec{v}_1\|^2}\vec{v}_1 = \frac{16}{13}\langle -2, 3\rangle = \left\langle -\frac{32}{13}, \frac{48}{13}\right\rangle$$

$$\text{proj}_{\vec{v}_2}\vec{x} = \frac{\vec{x}\cdot\vec{v}_2}{\|\vec{v}_2\|^2}\vec{v}_2 = -\frac{32}{52}\langle 4, -6\rangle = \left\langle -\frac{32}{13}, \frac{48}{13}\right\rangle$$

Exercise 7.2.2 Find the point on the line L defined by $4x - y = 0$ closest to the point $P = (6, 1)$. What is the distance between the point P and the line L?

Answer: The line L, which has equation $y = 4x$ is parallel to the vector $\vec{v} = \langle 1, 4\rangle$. Let $\vec{x} = \overrightarrow{OP} = \langle 6, 1\rangle$. Then

$$\text{proj}_{\vec{v}}\vec{x} = \frac{\vec{x}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v} = \frac{10}{17}\langle 1, 4\rangle = \left\langle \frac{10}{17}, \frac{40}{17}\right\rangle.$$

The point on L that is closest to P is $\left(\frac{10}{17}, \frac{40}{17}\right)$. The orthogonal part is

$$\vec{x} - \text{proj}_{\vec{v}}\vec{x} = \langle 6, 1\rangle - \left\langle \frac{10}{17}, \frac{40}{17}\right\rangle = \left\langle \frac{92}{17}, -\frac{23}{17}\right\rangle,$$

and the distance from P to L is

$$\|\vec{x} - \text{proj}_{\vec{v}}\vec{x}\| = \frac{23}{\sqrt{17}} \approx 5.58 \text{ (appropriate length units).}$$

Exercise 7.2.3

1. Let $\vec{v} = \langle 1, -1, 2, -3\rangle$ and $\vec{x} = \langle 3, -3, 6, -9\rangle$. Verify that $\vec{x}$ is parallel to $\vec{v}$ and that $\text{proj}_{\vec{v}}\vec{x} = \vec{x}$.
   Answer: A vector is parallel to $\vec{v}$ if it is a scalar multiple of $\vec{v}$. Since $\vec{x} = \langle 3, -3, 6, -9\rangle = 3\langle 1, -1, 2, -3\rangle = 3\vec{v}$, $\vec{x}$ is parallel to $\vec{v}$. Using the projection formula,

   $$\text{proj}_{\vec{v}}\vec{x} = \frac{\vec{x}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v} = \frac{45}{15}\langle 1, -1, 2, -3\rangle = \langle 3, -3, 6, -9\rangle = \vec{x},$$

   as expected.

2. Let $\vec{v}$ be any nonzero vector in R^n. Show that if $\vec{x}$ is any vector in R^n that is parallel to $\vec{v}$, then $\text{proj}_{\vec{v}}\,\vec{x} = \vec{x}$.
   Answer: Since $\vec{x}$ is parallel to $\vec{v}$, $\vec{x} = c\vec{v}$ for some scalar c. Then

$$\text{proj}_{\vec{v}}\,\vec{x} = \frac{\vec{x}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v} = \frac{c\vec{v}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v} = \frac{c\|\vec{v}\|^2}{\|\vec{v}\|^2}\vec{v} = c\vec{v} = \vec{x}.$$

   This is the anticipated result.

Chapter 7 Additional Exercises

Bibliography

[1] Strang, G., Introduction to Linear Algebra, 3rd Edition, Harcourt, Brace, Jovanovich, Publishers, 1988.

[2] Strang, G., The Fundamental Theorem of Linear Algebra, *The American Mathematical Monthly*, vol. 100, no. 9, 1993, pp. 848-855.

[3] Strang, G., Introduction to Linear Algebra, 6th Edition, Wellesley Cambridge Press, 2023