A FIRST COURSE
IN LINEAR ALGEBRA

An Open Text* by Ken Kuttler

LYRYX with OPEN TEXTS ADAPTATION

YORK UNIVERSITY

MATH 1021- LINEAR ALGEBRA I
FALL 2015- All Sections

ADAPTED OPEN TEXT

FORMATIVE ONLINE ASSESSMENT

COURSE SUPPLEMENTS

COURSE LOGISTICS & SUPPORT

lyryx
advancing learning
www.lyryx.com

A First Course in Linear Algebra

Ken Kuttler

Version 2015 Revision A

CONTENTS

Preface

Linear Algebra: A First Course presents an introduction to the fascinating subject of linear algebra. As the title suggests, this text is designed as a first course in linear algebra for students who have a reasonable understanding of basic algebra. Major topics of linear algebra are presented in detail, with proofs of important theorems provided. Connections to additional topics covered in advanced courses are introduced, in an effort to assist those students who are interested in continuing on in linear algebra.

Each chapter begins with a list of desired outcomes which a student should be able to achieve upon completing the chapter. Throughout the text, examples and diagrams are given to reinforce ideas and provide guidance on how to approach various problems. Suggested exercises are given at the end of each section, and students are encouraged to work through a selection of these exercises.

A brief review of complex numbers is given, which can serve as an introduction to anyone unfamiliar with the topic.

Linear algebra is a wonderful and interesting subject, which should not be limited to a challenge of correct arithmetic. The use of a computer algebra system can be a great help in long and difficult computations. Some of the standard computations of linear algebra are easily done by the computer, including finding the reduced row-echelon form. While the use of a computer system is encouraged, it is not meant to be done without the student having an understanding of the computations.

1. Systems of Equations

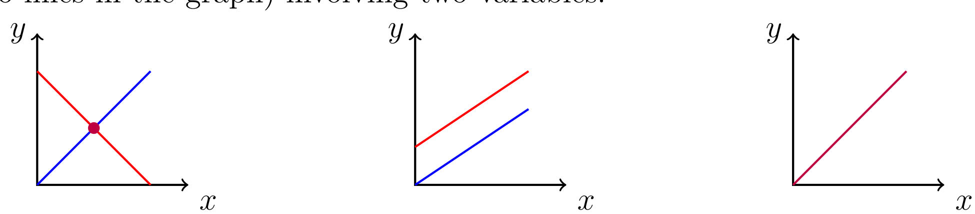
1.1 Systems of Equations, Geometry

Outcomes

A. *Relate the types of solution sets of a system of two (three) variables to the intersections of lines in a plane (the intersection of planes in three space)*

As you may remember, linear equations like $2x + 3y = 6$ can be graphed as straight lines in the coordinate plane. We say that this equation is in two variables, in this case x and y. Suppose you have two such equations, each of which can be graphed as a straight line, and consider the resulting graph of two lines. What would it mean if there exists a point of intersection between the two lines? This point, which lies on *both* graphs, gives x and y values for which both equations are true. In other words, this point gives the ordered pair (x, y) that satisfy both equations. If the point (x, y) is a point of intersection, we say that (x, y) is a **solution** to the two equations. In linear algebra, we often are concerned with finding the solution(s) to a system of equations, if such solutions exist. First, we consider graphical representations of solutions and later we will consider the algebraic methods for finding solutions.

When looking for the intersection of two lines in a graph, several situations may arise. The following picture demonstrates the possible situations when considering two equations (two lines in the graph) involving two variables.



One Solution   No Solutions   Infinitely Many Solutions

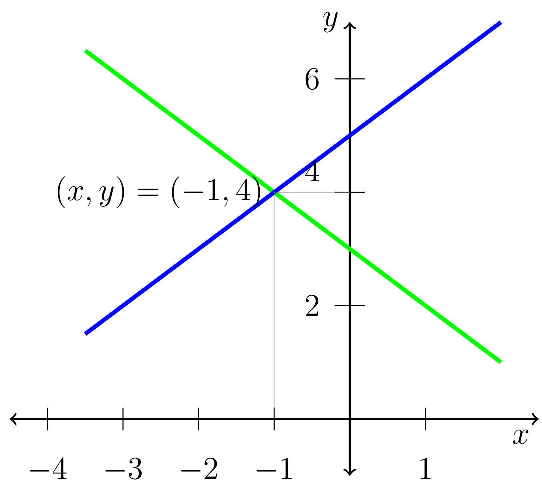
In the first diagram, there is a unique point of intersection, which means that there is only one (unique) solution to the two equations. In the second, there are no points of intersection and no solution. When no solution exists, this means that the two lines are parallel and they never intersect. The third situation which can occur, as demonstrated in diagram three, is that the two lines are really the same line. For example, $x + y = 1$ and $2x + 2y = 2$ are equations which when graphed yield the same line. In this case there are infinitely many points which are solutions of these two equations, as every ordered pair which is on the graph of the line satisfies both equations. When considering linear systems of equations, there are always three types of solutions possible; exactly one (unique) solution, infinitely many solutions, or no solution.

Example 1.1: A Graphical Solution

Use a graph to find the solution to the following system of equations

$$\begin{aligned} x + y &= 3 \\ y - x &= 5 \end{aligned}$$

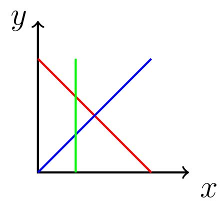
Solution. Through graphing the above equations and identifying the point of intersection, we can find the solution(s). Remember that we must have either one solution, infinitely many, or no solutions at all. The following graph shows the two equations, as well as the intersection. Remember, the point of intersection represents the solution of the two equations, or the (x, y) which satisfy both equations. In this case, there is one point of intersection at $(-1, 4)$ which means we have one unique solution, $x = -1, y = 4$.



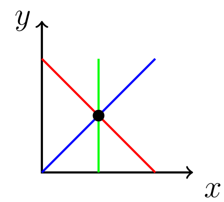
□

In the above example, we investigated the intersection point of two equations in two variables, x and y. Now we will consider the graphical solutions of three equations in two variables.

Consider a system of three equations in two variables. Again, these equations can be graphed as straight lines in the plane, so that the resulting graph contains three straight lines. Recall the three possible types of solutions; no solution, one solution, and infinitely many solutions. There are now more complex ways of achieving these situations, due to the presence of the third line. For example, you can imagine the case of three intersecting lines having no common point of intersection. Perhaps you can also imagine three intersecting lines which do intersect at a single point. These two situations are illustrated below.



No Solution

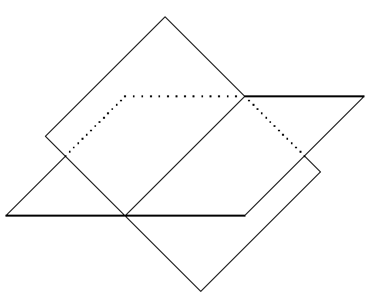


One Solution

Consider the first picture above. While all three lines intersect with one another, there is no common point of intersection where all three lines meet at one point. Hence, there is no solution to the three equations. Remember, a solution is a point (x, y) which satisfies **all** three equations. In the case of the second picture, the lines intersect at a common point. This means that there is one solution to the three equations whose graphs are the given lines. You should take a moment now to draw the graph of a system which results in three parallel lines. Next, try the graph of three identical lines. Which type of solution is represented in each of these graphs?

We have now considered the graphical solutions of systems of two equations in two variables, as well as three equations in two variables. However, there is no reason to limit our investigation to equations in two variables. We will now consider equations in three variables.

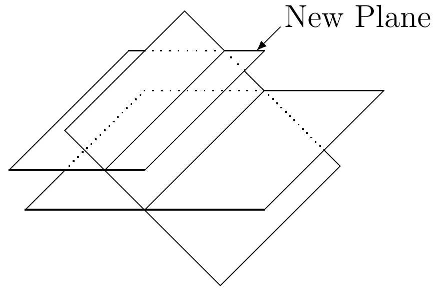
You may recall that equations in three variables, such as $2x + 4y - 5z = 8$, form a plane. Above, we were looking for intersections of lines in order to identify any possible solutions. When graphically solving systems of equations in three variables, we look for intersections of planes. These points of intersection give the (x, y, z) that satisfy all the equations in the system. What types of solutions are possible when working with three variables? Consider the following picture involving two planes, which are given by two equations in three variables.



Notice how these two planes intersect in a line. This means that the points (x, y, z) on this line satisfy both equations in the system. Since the line contains infinitely many points, this system has infinitely many solutions.

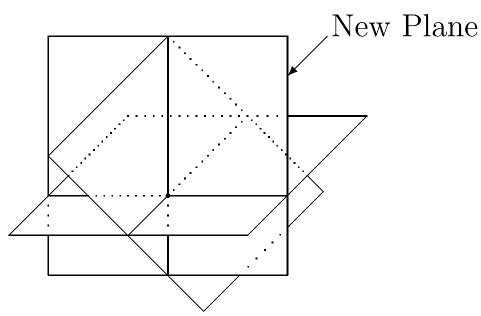
It could also happen that the two planes fail to intersect. However, is it possible to have two planes intersect at a single point? Take a moment to attempt drawing this situation, and convince yourself that it is not possible! This means that when we have only two equations in three variables, there is no way to have a unique solution! Hence, the types of solutions possible for two equations in three variables are no solution or infinitely many solutions.

Now imagine adding a third plane. In other words, consider three equations in three variables. What types of solutions are now possible? Consider the following diagram.

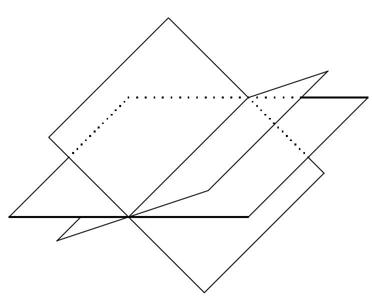


In this diagram, there is no point which lies in all three planes. There is no intersection between **all** planes so there is no solution. The picture illustrates the situation in which the line of intersection of the new plane with one of the original planes forms a line parallel to the line of intersection of the first two planes. However, in three dimensions, it is possible for two lines to fail to intersect even though they are not parallel. Such lines are called **skew lines.**

Recall that when working with two equations in three variables, it was not possible to have a unique solution. Is it possible when considering three equations in three variables? In fact, it is possible, and we demonstrate this situation in the following picture.



In this case, the three planes have a single point of intersection. Can you think of other types of solutions possible? Another is that the three planes could intersect in a line, resulting in infinitely many solutions, as in the following diagram.



We have now seen how three equations in three variables can have no solution, a unique solution, or intersect in a line resulting in infinitely many solutions. It is also possible that the three equations graph the same plane, which also leads to infinitely many solutions.

You can see that when working with equations in three variables, there are many more ways to achieve the different types of solutions than when working with two variables. It may prove enlightening to spend time imagining (and drawing) many possible scenarios, and you should take some time to try a few.

You should also take some time to imagine (and draw) graphs of systems in more than three variables. Equations like $x + y - 2z + 4w = 8$ with more than three variables are often called **hyper-planes**. You may soon realize that it is tricky to draw the graphs of hyper-planes! Through the tools of linear algebra, we can algebraically examine these types of systems which are difficult to graph. In the following section, we will consider these algebraic tools.

Exercises

Exercise 1.1.1 *Graphically, find the point* (x_1, y_1) *which lies on both lines,* $x + 3y = 1$ *and* $4x - y = 3$*. That is, graph each line and see where they intersect.*

Exercise 1.1.2 *Graphically, find the point of intersection of the two lines* $3x + y = 3$ *and* $x + 2y = 1$*. That is, graph each line and see where they intersect.*

Exercise 1.1.3 *You have a system of* k *equations in two variables,* $k \geq 2$*. Explain the geometric significance of*

1. *No solution.*
2. *A unique solution.*
3. *An infinite number of solutions.*

1.2 Systems Of Equations, Algebraic Procedures

Outcomes

A. *Use elementary operations to find the solution to a linear system of equations.*

B. *Find the row-echelon form and reduced row-echelon form of a matrix.*

C. *Determine whether a system of linear equations has no solution, a unique solution or an infinite number of solutions from its row-echelon form.*

D. *Solve a system of equations using Gaussian Elimination and Gauss-Jordan Elimination.*

E. *Model a physical system with linear equations and then solve.*

We have taken an in depth look at graphical representations of systems of equations, as well as how to find possible solutions graphically. Our attention now turns to working with systems algebraically.

Definition 1.2: System of Linear Equations

A ***system of linear equations*** *is a list of equations,*

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

where a_{ij} and b_j are real numbers. The above is a system of m equations in the n variables, $x_1, x_2 \cdots, x_n$. Written more simply in terms of summation notation, the above can be written in the form

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \ i = 1, 2, 3, \cdots, m$$

The relative size of m and n is not important here. Notice that we have allowed a_{ij} and b_j to be any real number. We can also call these numbers **scalars** . We will use this term throughout the text, so keep in mind that the term **scalar** just means that we are working with real numbers.

Now, suppose we have a system where $b_i = 0$ for all i. In other words every equation equals 0. This is a special type of system.

Definition 1.3: Homogeneous System of Equations

A system of equations is called ***homogeneous*** *if each equation in the system is equal to* 0*. A homogeneous system has the form*

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

where a_{ij} *are scalars and* x_i *are variables.*

Recall from the previous section that our goal when working with systems of linear equations was to find the point of intersection of the equations when graphed. In other words, we looked for the solutions to the system. We now wish to find these solutions algebraically. We want to find values for $x_1, \cdots, x_n$ which solve all of the equations. If such a set of values exists, we call $(x_1, \cdots, x_n)$ the **solution set**.

Recall the above discussions about the types of solutions possible. We will see that systems of linear equations will have one unique solution, infinitely many solutions, or no solution. Consider the following definition.

Definition 1.4: Consistent and Inconsistent Systems

A system of linear equations is called ***consistent*** *if there exists at least one solution. It is called* ***inconsistent*** *if there is no solution.*

If you think of each equation as a condition which must be satisfied by the variables, consistent would mean there is some choice of variables which can satisfy **all** the conditions. Inconsistent would mean there is no choice of the variables which can satisfy all of the conditions.

The following sections provide methods for determining if a system is consistent or inconsistent, and finding solutions if they exist.

1.2.1. Elementary Operations

We begin this section with an example. Recall from Example 1.1 that the solution to the given system was $(x, y) = (-1, 4)$.

Example 1.5: Verifying an Ordered Pair is a Solution

Algebraically verify that $(x, y) = (-1, 4)$ *is a solution to the following system of equations.*

$$\begin{aligned}
x + y &= 3 \\
y - x &= 5
\end{aligned}$$

Solution. By graphing these two equations and identifying the point of intersection, we previously found that $(x, y) = (-1, 4)$ is the unique solution.

We can verify algebraically by substituting these values into the original equations, and ensuring that the equations hold. First, we substitute the values into the first equation and check that it equals 3.

$$x + y = (-1) + (4) = 3$$

This equals 3 as needed, so we see that $(-1, 4)$ is a solution to the first equation. Substituting the values into the second equation yields

$$y - x = (4) - (-1) = 4 + 1 = 5$$

which is true. For $(x, y) = (-1, 4)$ each equation is true and therefore, this is a solution to the system. ◻

Now, the interesting question is this: If you were not given these numbers to verify, how could you algebraically determine the solution? Linear algebra gives us the tools needed to answer this question. The following basic operations are important tools that we will utilize.

Definition 1.6: Elementary Operations

Elementary operations *are those operations consisting of the following.*

1. *Interchange the order in which the equations are listed.*
2. *Multiply any equation by a nonzero number.*
3. *Replace any equation with itself added to a multiple of another equation.*

It is important to note that none of these operations will change the set of solutions of the system of equations. In fact, elementary operations are the *key tool* we use in linear algebra to find solutions to systems of equations.

Consider the following example.

Example 1.7: Effects of an Elementary Operation

Show that the system

$$\begin{aligned} x + y &= 7 \\ 2x - y &= 8 \end{aligned}$$

has the same solution as the system

$$\begin{aligned} x + y &= 7 \\ -3y &= -6 \end{aligned}$$

Solution. Notice that the second system has been obtained by taking the second equation of the first system and adding -2 times the first equation, as follows:

$$2x - y + (-2)(x + y) = 8 + (-2)(7)$$

By simplifying, we obtain

$$-3y = -6$$

which is the second equation in the second system. Now, from here we can solve for y and see that $y = 2$. Next, we substitute this value into the first equation as follows

$$x + y = x + 2 = 7$$

Hence $x = 5$ and so $(x, y) = (5, 2)$ is a solution to the second system. We want to check if $(5, 2)$ is also a solution to the first system. We check this by substituting $(x, y) = (5, 2)$ into the system and ensuring the equations are true.

$$\begin{aligned} x + y &= (5) + (2) = 7 \\ 2x - y &= 2(5) - (2) = 8 \end{aligned}$$

Hence, $(5, 2)$ is also a solution to the first system. ☐

This example illustrates how an elementary operation applied to a system of two equations in two variables does not affect the solution set. However, a linear system may involve many equations and many variables and there is no reason to limit our study to small systems. For any size of system in any number of variables, the solution set is still the collection of solutions to the equations. In every case, the above operations of Definition 1.6 do not change the set of solutions to the system of linear equations.

In the following theorem, we use the notation E_i to represent an equation, while b_i denotes a constant.

Theorem 1.8: Elementary Operations and Solutions

Suppose you have a system of two linear equations

$$\begin{aligned} E_1 &= b_1 \\ E_2 &= b_2 \end{aligned} \tag{1.1}$$

Then the following systems have the same solution set as 1.1:

1. $$\begin{aligned} E_2 &= b_2 \\ E_1 &= b_1 \end{aligned} \tag{1.2}$$

2. $$\begin{aligned} E_1 &= b_1 \\ kE_2 &= kb_2 \end{aligned} \tag{1.3}$$

   for any scalar k, provided $k \neq 0$.

3. $$\begin{aligned} E_1 &= b_1 \\ E_2 + kE_1 &= b_2 + kb_1 \end{aligned} \tag{1.4}$$

   for any scalar k (including $k = 0$).

Before we proceed with the proof of Theorem 1.8, let us consider this theorem in context of Example 1.7. Then,

$$\begin{aligned} E_1 &= x + y, \quad b_1 = 7 \\ E_2 &= 2x - y, \quad b_2 = 8 \end{aligned}$$

Recall the elementary operations that we used to modify the system in the solution to the example. First, we added (-2) times the first equation to the second equation. In terms of Theorem 1.8, this action is given by

$$E_2 + (-2)\,E_1 = b_2 + (-2)\,b_1$$

or

$$2x - y + (-2)\,(x + y) = 8 + (-2)\,7$$

This gave us the second system in Example 1.7, given by

$$\begin{aligned} E_1 &= b_1 \\ E_2 + (-2)\,E_1 &= b_2 + (-2)\,b_1 \end{aligned}$$

From this point, we were able to find the solution to the system. Theorem 1.8 tells us that the solution we found is in fact a solution to the original system.

We will now prove Theorem 1.8.

Proof.

1. The proof that the systems 1.1 and 1.2 have the same solution set is as follows. Suppose that $(x_1, \cdots, x_n)$ is a solution to $E_1 = b_1, E_2 = b_2$. We want to show that this is a solution to the system in 1.2 above. This is clear, because the system in 1.2 is the original system, but listed in a different order. Changing the order does not effect the solution set, so $(x_1, \cdots, x_n)$ is a solution to 1.2.

2. Next we want to prove that the systems 1.1 and 1.3 have the same solution set. That is $E_1 = b_1, E_2 = b_2$ has the same solution set as the system $E_1 = b_1, kE_2 = kb_2$ provided $k \neq 0$. Let $(x_1, \cdots, x_n)$ be a solution of $E_1 = b_1, E_2 = b_2,$. We want to show that it is a solution to $E_1 = b_1, kE_2 = kb_2$. Notice that the only difference between these two systems is that the second involves multiplying the equation, $E_2 = b_2$ by the scalar k. Recall that when you multiply both sides of an equation by the same number, the sides are still equal to each other. Hence if $(x_1, \cdots, x_n)$ is a solution to $E_2 = b_2$, then it will also be a solution to $kE_2 = kb_2$. Hence, $(x_1, \cdots, x_n)$ is also a solution to 1.3.

   Similarly, let $(x_1, \cdots, x_n)$ be a solution of $E_1 = b_1, kE_2 = kb_2$. Then we can multiply the equation $kE_2 = kb_2$ by the scalar $1/k$, which is possible only because we have required that $k \neq 0$. Just as above, this action preserves equality and we obtain the equation $E_2 = b_2$. Hence $(x_1, \cdots, x_n)$ is also a solution to $E_1 = b_1, E_2 = b_2$.

3. Finally, we will prove that the systems 1.1 and 1.4 have the same solution set. We will show that any solution of $E_1 = b_1, E_2 = b_2$ is also a solution of 1.4. Then, we will show that any solution of 1.4 is also a solution of $E_1 = b_1, E_2 = b_2$. Let $(x_1, \cdots, x_n)$ be a solution to $E_1 = b_1, E_2 = b_2$. Then in particular it solves $E_1 = b_1$. Hence, it solves the

first equation in 1.4. Similarly, it also solves $E_2 = b_2$. By our proof of 1.3, it also solves $kE_1 = kb_1$. Notice that if we add E_2 and kE_1, this is equal to $b_2 + kb_1$. Therefore, if $(x_1, \cdots, x_n)$ solves $E_1 = b_1, E_2 = b_2$ it must also solve $E_2 + kE_1 = b_2 + kb_1$.

Now suppose $(x_1, \cdots, x_n)$ solves the system $E_1 = b_1, E_2 + kE_1 = b_2 + kb_1$. Then in particular it is a solution of $E_1 = b_1$. Again by our proof of 1.3, it is also a solution to $kE_1 = kb_1$. Now if we subtract these equal quantities from both sides of $E_2 + kE_1 = b_2 + kb_1$ we obtain $E_2 = b_2$, which shows that the solution also satisfies $E_1 = b_1, E_2 = b_2$.

□

Stated simply, the above theorem shows that the elementary operations do not change the solution set of a system of equations.

We will now look at an example of a system of three equations and three variables. Similarly to the previous examples, the goal is to find values for x, y, z such that each of the given equations are satisfied when these values are substituted in.

Example 1.9: Solving a System of Equations with Elementary Operations

Find the solutions to the system,

$$\begin{array}{c} x + 3y + 6z = 25 \\ 2x + 7y + 14z = 58 \\ 2y + 5z = 19 \end{array} \tag{1.5}$$

Solution. We can relate this system to Theorem 1.8 above. In this case, we have

$$\begin{array}{ll} E_1 = x + 3y + 6z, & b_1 = 25 \\ E_2 = 2x + 7y + 14z, & b_2 = 58 \\ E_3 = 2y + 5z, & b_3 = 19 \end{array}$$

Theorem 1.8 claims that if we do elementary operations on this system, we will not change the solution set. Therefore, we can solve this system using the elementary operations given in Definition 1.6. First, replace the second equation by (-2) times the first equation added to the second. This yields the system

$$\begin{array}{c} x + 3y + 6z = 25 \\ y + 2z = 8 \\ 2y + 5z = 19 \end{array} \tag{1.6}$$

Now, replace the third equation with (-2) times the second added to the third. This yields the system

$$\begin{array}{c} x + 3y + 6z = 25 \\ y + 2z = 8 \\ z = 3 \end{array} \tag{1.7}$$

At this point, we can easily find the solution. Simply take $z = 3$ and substitute this back into the previous equation to solve for y, and similarly to solve for x.

$$\begin{aligned} x + 3y + 6(3) = x + 3y + 18 &= 25 \\ y + 2(3) = y + 6 &= 8 \\ z &= 3 \end{aligned}$$

The second equation is now

$$y + 6 = 8$$

You can see from this equation that $y = 2$. Therefore, we can substitute this value into the first equation as follows:

$$x + 3(2) + 18 = 25$$

By simplifying this equation, we find that $x = 1$. Hence, the solution to this system is $(x, y, z) = (1, 2, 3)$. This process is called **back substitution**.

Alternatively, in 1.7 you could have continued as follows. Add (-2) times the third equation to the second and then add (-6) times the second to the first. This yields

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

Now add (-3) times the second to the first. This yields

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

a system which has the same solution set as the original system. This avoided back substitution and led to the same solution set. It is your decision which you prefer to use, as both methods lead to the correct solution, $(x, y, z) = (1, 2, 3)$. ∎

1.2.2. Gaussian Elimination

The work we did in the previous section will always find the solution to the system. In this section, we will explore a less cumbersome way to find the solutions. First, we will represent a linear system with an **augmented matrix.** A **matrix** is simply a rectangular array of numbers. The size or dimension of a matrix is defined as $m \times n$ where m is the number of rows and n is the number of columns. In order to construct an augmented matrix from a linear system, we create a **coefficient matrix** from the coefficients of the variables in the system, as well as a **constant matrix** from the constants. The coefficients from one equation of the system create one row of the augmented matrix.

For example, consider the linear system in Example 1.9

$$\begin{aligned} x + 3y + 6z &= 25 \\ 2x + 7y + 14z &= 58 \\ 2y + 5z &= 19 \end{aligned}$$

This system can be written as an augmented matrix, as follows

$$\left[\begin{array}{rrr|r} 1 & 3 & 6 & 25 \\ 2 & 7 & 14 & 58 \\ 0 & 2 & 5 & 19 \end{array}\right]$$

Notice that it has exactly the same information as the original system. Here it is understood that the first column contains the coefficients from x in each equation, in order, $\left[\begin{array}{c} 1 \\ 2 \\ 0 \end{array}\right]$. Similarly, we create a column from the coefficients on y in each equation, $\left[\begin{array}{c} 3 \\ 7 \\ 2 \end{array}\right]$ and a column from the coefficients on z in each equation, $\left[\begin{array}{r} 6 \\ 14 \\ 5 \end{array}\right]$. For a system of more than three variables, we would continue in this way constructing a column for each variable. Similarly, for a system of less than three variables, we simply construct a column for each variable.

Finally, we construct a column from the constants of the equations, $\left[\begin{array}{c} 25 \\ 58 \\ 19 \end{array}\right]$.

The rows of the augmented matrix correspond to the equations in the system. For example, the top row in the augmented matrix, $\left[\begin{array}{ccc|c} 1 & 3 & 6 & 25 \end{array}\right]$ corresponds to the equation

$$x + 3y + 6z = 25.$$

Consider the following definition.

Definition 1.10: Augmented Matrix of a Linear System

For a linear system of the form

$$\begin{array}{c} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{array}$$

where the x_i are variables and the a_{ij} and b_i are constants, the augmented matrix of this system is given by

$$\left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right]$$

Now, consider elementary operations in the context of the augmented matrix. The elementary operations in Definition 1.6 can be used on the rows just as we used them on equations previously. Changes to a system of equations in as a result of an elementary operation are equivalent to changes in the augmented matrix resulting from the corresponding row operation. Note that Theorem 1.8 implies that any elementary row operations used on an augmented matrix will not change the solution to the corresponding system of equations. We now formally define elementary row operations. These are the *key tool* we will use to find solutions to systems of equations.

Definition 1.11: Elementary Row Operations

*The **elementary row operations** (also known as **row operations**) consist of the following*

1. *Switch two rows.*
2. *Multiply a row by a nonzero number.*
3. *Replace a row by any multiple of another row added to it.*

Recall how we solved Example 1.9. We can do the exact same steps as above, except now in the context of an augmented matrix and using row operations. The augmented matrix of this system is

$$\left[\begin{array}{rrr|r} 1 & 3 & 6 & 25 \\ 2 & 7 & 14 & 58 \\ 0 & 2 & 5 & 19 \end{array}\right]$$

Thus the first step in solving the system given by 1.5 would be to take (-2) times the first row of the augmented matrix and add it to the second row,

$$\left[\begin{array}{rrr|r} 1 & 3 & 6 & 25 \\ 0 & 1 & 2 & 8 \\ 0 & 2 & 5 & 19 \end{array}\right]$$

Note how this corresponds to 1.6. Next take (-2) times the second row and add to the third,

$$\left[\begin{array}{rrr|r} 1 & 3 & 6 & 25 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

This augmented matrix corresponds to the system

$$\begin{array}{c} x + 3y + 6z = 25 \\ y + 2z = 8 \\ z = 3 \end{array}$$

which is the same as 1.7. By back substitution you obtain the solution $x = 1, y = 2$, and $z = 3$.

Through a systematic procedure of row operations, we can simplify an augmented matrix and carry it to **row-echelon form** or **reduced row-echelon form**, which we define next. These forms are used to find the solutions of the system of equations corresponding to the augmented matrix.

In the following definitions, the term **leading entry** refers to the first nonzero entry of a row when scanning the row from left to right.

Definition 1.12: Row-Echelon Form

An augmented matrix is in ***row-echelon form*** *if*

1. *All nonzero rows are above any rows of zeros.*
2. *Each leading entry of a row is in a column to the right of the leading entries of any row above it.*
3. *Each leading entry of a row is equal to* 1*.*

We also consider another reduced form of the augmented matrix which has one further condition.

Definition 1.13: Reduced Row-Echelon Form

An augmented matrix is in ***reduced row-echelon form*** *if*

1. *All nonzero rows are above any rows of zeros.*
2. *Each leading entry of a row is in a column to the right of the leading entries of any rows above it.*
3. *Each leading entry of a row is equal to* 1*.*
4. *All entries in a column above and below a leading entry are zero.*

Notice that the first three conditions on a reduced row-echelon form matrix are the same as those for row-echelon form.

Hence, every reduced row-echelon form matrix is also in row-echelon form. The converse is not necessarily true; we cannot assume that every matrix in row-echelon form is also in reduced row-echelon form. However, it often happens that the row-echelon form is sufficient to provide information about the solution of a system.

The following examples describe matrices in these various forms. As an exercise, take the time to carefully verify that they are in the specified form.

Example 1.14: Not in Row-Echelon Form

The following augmented matrices are not in row-echelon form (and therefore also not in reduced row-echelon form).

$$\left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right], \left[\begin{array}{cc|r} 1 & 2 & 3 \\ 2 & 4 & -6 \\ 4 & 0 & 7 \end{array}\right], \left[\begin{array}{ccc|c} 0 & 2 & 3 & 3 \\ 1 & 5 & 0 & 2 \\ 7 & 5 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

Example 1.15: Matrices in Row-Echelon Form

The following augmented matrices are in row-echelon form, but not in reduced row-echelon form.

$$\left[\begin{array}{ccccc|c} 1 & 0 & 6 & 5 & 8 & 2 \\ 0 & 0 & 1 & 2 & 7 & 3 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right], \left[\begin{array}{ccc|c} 1 & 3 & 5 & 4 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right], \left[\begin{array}{ccc|c} 1 & 0 & 6 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Notice that we could apply further row operations to these matrices to carry them to reduced row-echelon form. Take the time to try that on your own. Consider the following matrices, which are in reduced row-echelon form.

Example 1.16: Matrices in Reduced Row-Echelon Form

The following augmented matrices are in reduced row-echelon form.

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

One way in which the row-echelon form of a matrix is useful is in identifying the pivot positions and pivot columns of the matrix.

Definition 1.17: Pivot Position and Pivot Column

*A **pivot position** in a matrix is the location of a leading entry in the row-echelon formof a matrix. A **pivot column** is a column that contains a pivot position.*

For example consider the following.

Example 1.18: Pivot Position

Let

$$A = \left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 6 \\ 4 & 4 & 4 & 10 \end{array}\right]$$

Where are the pivot positions and pivot columns of the augmented matrix A?

Solution. The row-echelon form of this matrix is

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

This is all we need in this example, but note that this matrix is not in reduced row-echelon form.

In order to identify the pivot positions in the original matrix, we look for the leading entries in the row-echelon form of the matrix. Here, the entry in the first row and first column, as well as the entry in the second row and second column are the leading entries. Hence, these locations are the pivot positions. We identify the pivot positions in the original matrix, as in the following:

$$\left[\begin{array}{rrr|r} \boxed{1} & 2 & 3 & 4 \\ 3 & \boxed{2} & 1 & 6 \\ 4 & 4 & 4 & 10 \end{array}\right]$$

Thus the pivot columns in the matrix are the first two columns. ▯

The following is an algorithm for carrying a matrix to row-echelon form and reduced row-echelon form. You may wish to use this algorithm to carry the above matrix to row-echelon form or reduced row-echelon form yourself for practice.

Algorithm 1.19: Reduced Row-Echelon Form Algorithm

This algorithm provides a method for using row operations to take a matrix to its reduced row-echelon form. We begin with the matrix in its original form.

1. *Starting from the left, find the first nonzero column. This is the first pivot column, and the position at the top of this column is the first pivot position. Switch rows if necessary to place a nonzero number in the first pivot position.*

2. *Use row operations to make the entries below the first pivot position (in the first pivot column) equal to zero.*

3. *Ignoring the row containing the first pivot position, repeat steps 1 and 2 with the remaining rows. Repeat the process until there are no more rows to modify.*

4. *Divide each nonzero row by the value of the leading entry, so that the leading entry becomes 1. The matrix will then be in row-echelon form.*

The following step will carry the matrix from row-echelon form to reduced row-echelon form.

5. *Moving from right to left, use row operations to create zeros in the entries of the pivot columns which are above the pivot positions. The result will be a matrix in reduced row-echelon form.*

Most often we will apply this algorithm to an augmented matrix in order to find the solution to a system of linear equations. However, we can use this algorithm to compute the reduced row-echelon form of any matrix which could be useful in other applications.

Consider the following example of Algorithm 1.19.

Example 1.20: Finding Row-Echelon Form and Reduced Row-Echelon Form of a Matrix

Let

$$A = \begin{bmatrix} 0 & -5 & -4 \\ 1 & 4 & 3 \\ 5 & 10 & 7 \end{bmatrix}$$

Find the row-echelon form of A. Then complete the process until A is in reduced row-echelon form.

Solution. In working through this example, we will use the steps outlined in Algorithm 1.19.

1. The first pivot column is the first column of the matrix, as this is the first nonzero column from the left. Hence the first pivot position is the one in the first row and first column. Switch the first two rows to obtain a nonzero entry in the first pivot position, outlined in a box below.
$$\begin{bmatrix} \boxed{1} & 4 & 3 \\ 0 & -5 & -4 \\ 5 & 10 & 7 \end{bmatrix}$$

2. Step two involves creating zeros in the entries below the first pivot position. The first entry of the second row is already a zero. All we need to do is subtract 5 times the first row from the third row. The resulting matrix is
$$\begin{bmatrix} 1 & 4 & 3 \\ 0 & -5 & -4 \\ 0 & 10 & 8 \end{bmatrix}$$

3. Now ignore the top row. Apply steps 1 and 2 to the smaller matrix
$$\begin{bmatrix} -5 & -4 \\ 10 & 8 \end{bmatrix}$$
In this matrix, the first column is a pivot column, and -5 is in the first pivot position. Therefore, we need to create a zero below it. To do this, add 2 times the first row (of this matrix) to the second. The resulting matrix is
$$\begin{bmatrix} -5 & -4 \\ 0 & 0 \end{bmatrix}$$
Our original matrix now looks like
$$\begin{bmatrix} 1 & 4 & 3 \\ 0 & -5 & -4 \\ 0 & 0 & 0 \end{bmatrix}$$
We can see that there are no more rows to modify.

4. Now, we need to create leading 1s in each row. The first row already has a leading 1 so no work is needed here. Divide the second row by -5 to create a leading 1. The resulting matrix is
$$\left[\begin{array}{ccc} 1 & 4 & 3 \\ 0 & 1 & \frac{4}{5} \\ 0 & 0 & 0 \end{array}\right]$$
This matrix is now in row-echelon form.

5. Now create zeros in the entries above pivot positions in each column, in order to carry this matrix all the way to reduced row-echelon form. Notice that there is no pivot position in the third column so we do not need to create any zeros in this column! The column in which we need to create zeros is the second. To do so, subtract 4 times the second row from the first row. The resulting matrix is
$$\left[\begin{array}{ccr} 1 & 0 & -\frac{1}{5} \\ 0 & 1 & \frac{4}{5} \\ 0 & 0 & 0 \end{array}\right]$$

This matrix is now in reduced row-echelon form. ∎

The above algorithm gives you a simple way to obtain the row-echelon form and reduced row-echelon form of a matrix. The main idea is to do row operations in such a way as to end up with a matrix in row-echelon form or reduced row-echelon form. This process is important because the resulting matrix will allow you to describe the solutions to the corresponding linear system of equations in a meaningful way.

In the next example, we look at how to solve a system of equations using the corresponding augmented matrix.

Example 1.21: Finding the Solution to a System

Give the complete solution to the following system of equations
$$\begin{array}{c} 2x+4y-3z=-1 \\ 5x+10y-7z=-2 \\ 3x+6y+5z=9 \end{array}$$

Solution. The augmented matrix for this system is
$$\left[\begin{array}{rrr|r} 2 & 4 & -3 & -1 \\ 5 & 10 & -7 & -2 \\ 3 & 6 & 5 & 9 \end{array}\right]$$

In order to find the solution to this system, we wish to carry the augmented matrix to reduced row-echelon form. We will do so using Algorithm 1.19. Notice that the first column is nonzero, so this is our first pivot column. The first entry in the first row, 2, is the first leading entry and it is in the first pivot position. We will use row operations to create zeros in the entries below the 2. First, replace the second row with -5 times the first row plus 2 times the second row. This yields

$$\left[\begin{array}{rrr|r} 2 & 4 & -3 & -1 \\ 0 & 0 & 1 & 1 \\ 3 & 6 & 5 & 9 \end{array}\right]$$

Now, replace the third row with -3 times the first row plus to 2 times the third row. This yields

$$\left[\begin{array}{rrr|r} 2 & 4 & -3 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 21 \end{array}\right]$$

Now the entries in the first column below the pivot position are zeros. We now look for the second pivot column, which in this case is column three. Here, the 1 in the second row and third column is in the pivot position. We need to do just one row operation to create a zero below the 1.

Taking -1 times the second row and adding it to the third row yields

$$\left[\begin{array}{rrr|r} 2 & 4 & -3 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 20 \end{array}\right]$$

We could proceed with the algorithm to carry this matrix to row-echelon form or reduced row-echelon form. However, remember that we are looking for the solutions to the system of equations. Take another look at the third row of the matrix. Notice that it corresponds to the equation

$$0x + 0y + 0z = 20$$

There is no solution to this equation because for all x, y, z, the left side will equal 0 and $0 \neq 20$. This shows there is no solution to the given system of equations. In other words, this system is inconsistent. ☐

The following is another example of how to find the solution to a system of equations by carrying the corresponding augmented matrix to reduced row-echelon form.

Example 1.22: An Infinite Set of Solutions

Give the complete solution to the system of equations

$$\begin{array}{c} 3x - y - 5z = 9 \\ y - 10z = 0 \\ -2x + y = -6 \end{array} \tag{1.8}$$

Solution. The augmented matrix of this system is

$$\left[\begin{array}{rrr|r} 3 & -1 & -5 & 9 \\ 0 & 1 & -10 & 0 \\ -2 & 1 & 0 & -6 \end{array}\right]$$

In order to find the solution to this system, we will carry the augmented matrix to reduced row-echelon form, using Algorithm 1.19. The first column is the first pivot column. We want to use row operations to create zeros beneath the first entry in this column, which is in the first pivot position. Replace the third row with 2 times the first row added to 3 times the third row. This gives

$$\left[\begin{array}{rrr|r} 3 & -1 & -5 & 9 \\ 0 & 1 & -10 & 0 \\ 0 & 1 & -10 & 0 \end{array}\right]$$

Now, we have created zeros beneath the 3 in the first column, so we move on to the second pivot column (which is the second column) and repeat the procedure. Take -1 times the second row and add to the third row.

$$\left[\begin{array}{rrr|r} 3 & -1 & -5 & 9 \\ 0 & 1 & -10 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

The entry below the pivot position in the second column is now a zero. Notice that we have no more pivot columns because we have only two leading entries.

At this stage, we also want the leading entries to be equal to one. To do so, divide the first row by 3.

$$\left[\begin{array}{rrr|r} 1 & -\frac{1}{3} & -\frac{5}{3} & 3 \\ 0 & 1 & -10 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

This matrix is now in row-echelon form.

Let's continue with row operations until the matrix is in reduced row-echelon form. This involves creating zeros above the pivot positions in each pivot column. This requires only one step, which is to add $\frac{1}{3}$ times the second row to the first row.

$$\left[\begin{array}{rrr|r} 1 & 0 & -5 & 3 \\ 0 & 1 & -10 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

This is in reduced row-echelon form, which you should verify using Definition 1.13. The equations corresponding to this reduced row-echelon form are

$$\begin{aligned} x - 5z &= 3 \\ y - 10z &= 0 \end{aligned}$$

or

$$\begin{aligned} x &= 3 + 5z \\ y &= 10z \end{aligned}$$

Observe that z is not restrained by any equation. In fact, z can equal any number. For example, we can let $z = t$, where we can choose t to be any number. In this context t is called a **parameter** . Therefore, the solution set of this system is

$$\begin{aligned} x &= 3 + 5t \\ y &= 10t \\ z &= t \end{aligned}$$

where t is arbitrary. The system has an infinite set of solutions which are given by these equations. For any value of t we select, x, y, and z will be given by the above equations. For example, if we choose $t = 4$ then the corresponding solution would be

$$\begin{aligned} x &= 3 + 5(4) = 23 \\ y &= 10(4) = 40 \\ z &= 4 \end{aligned}$$

□

In Example 1.22 the solution involved one parameter. It may happen that the solution to a system involves more than one parameter, as shown in the following example.

Example 1.23: A Two Parameter Set of Solutions

Find the solution to the system

$$\begin{aligned} x + 2y - z + w &= 3 \\ x + y - z + w &= 1 \\ x + 3y - z + w &= 5 \end{aligned}$$

Solution. The augmented matrix is

$$\left[\begin{array}{rrrr|r} 1 & 2 & -1 & 1 & 3 \\ 1 & 1 & -1 & 1 & 1 \\ 1 & 3 & -1 & 1 & 5 \end{array}\right]$$

We wish to carry this matrix to row-echelon form. Here, we will outline the row operations used. However, make sure that you understand the steps in terms of Algorithm 1.19.

Take -1 times the first row and add to the second. Then take -1 times the first row and add to the third. This yields

$$\left[\begin{array}{rrrr|r} 1 & 2 & -1 & 1 & 3 \\ 0 & -1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 0 & 2 \end{array}\right]$$

Now add the second row to the third row and divide the second row by -1.

$$\left[\begin{array}{rrrr|r} 1 & 2 & -1 & 1 & 3 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{1.9}$$

This matrix is in row-echelon form and we can see that x and y correspond to pivot columns, while z and w do not. Therefore, we will assign parameters to the variables z and w. Assign the parameter s to z and the parameter t to w. Then the first row yields the equation $x + 2y - s + t = 3$, while the second row yields the equation $y = 2$. Since $y = 2$, the first equation becomes $x + 4 - s + t = 3$ showing that the solution is given by

$$\begin{aligned} x &= -1 + s - t \\ y &= 2 \\ z &= s \\ w &= t \end{aligned}$$

It is customary to write this solution in the form

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} -1+s-t \\ 2 \\ s \\ t \end{bmatrix} \tag{1.10}$$

$\square$

This example shows a system of equations with an infinite solution set which depends on two parameters. It can be less confusing in the case of an infinite solution set to first place the augmented matrix in reduced row-echelon form rather than just row-echelon form before seeking to write down the description of the solution.

In the above steps, this means we don't stop with the row-echelon form in equation 1.9. Instead we first place it in reduced row-echelon form as follows.

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

Then the solution is $y = 2$ from the second row and $x = -1 + z - w$ from the first. Thus letting $z = s$ and $w = t$, the solution is given by 1.10.

You can see here that there are two paths to the correct answer, which both yield the same answer. Hence, either approach may be used. The process which we first used in the above solution is called **Gaussian Elimination** This process involves carrying the matrix to row-echelon form, converting back to equations, and using back substitution to find the solution. When you do row operations until you obtain reduced row-echelon form, the process is called **Gauss-Jordan Elimination**.

We have now found solutions for systems of equations with no solution and infinitely many solutions, with one parameter as well as two parameters. Recall the three types of solution sets which we discussed in the previous section; no solution, one solution, and infinitely many solutions. Each of these types of solutions could be identified from the graph of the system. It turns out that we can also identify the type of solution from the reduced row-echelon form of the augmented matrix.

- *No Solution:* In the case where the system of equations has no solution, the row-echelon form of the augmented matrix will have a row of the form

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

  This row indicates that the system is inconsistent and has no solution.

- *One Solution:* In the case where the system of equations has one solution, every column of the coefficient matrix is a pivot column. The following is an example of an augmented matrix in reduced row-echelon form for a system of equations with one solution.

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

- *Infinitely Many Solutions:* In the case where the system of equations has infinitely many solutions, the solution contains parameters. There will be columns of the coefficient matrix which are not pivot columns. The following are examples of augmented matrices in reduced row-echelon form for systems of equations with infinitely many solutions.

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

or

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

1.2.3. Uniqueness of the Reduced Row-Echelon Form

As we have seen in earlier sections, we know that every matrix can be brought into reduced row-echelon form by a sequence of elementary row operations. Here we will prove that the resulting matrix is unique; in other words, the resulting matrix in reduced row-echelon form does not depend upon the particular sequence of elementary row operations or the order in which they were performed.

Let A be the augmented matrix of a homogeneous system of linear equations in the variables $x_1, x_2, \cdots, x_n$ which is also in reduced row-echelon form. The matrix A divides the set of variables in two different types. We say that x_i is a *basic variable* whenever A has a leading 1 in column number i, in other words, when column i is a pivot column. Otherwise we say that x_i is a *free variable*.

Recall Example 1.23.

Example 1.24: Basic and Free Variables

Find the basic and free variables in the system

$$\begin{gathered} x + 2y - z + w = 3 \\ x + y - z + w = 1 \\ x + 3y - z + w = 5 \end{gathered}$$

Solution. Recall from the solution of Example 1.23 that the row-echelon form of the augmented matrix of this system is given by

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

You can see that columns 1 and 2 are pivot columns. These columns correspond to variables x and y, making these the basic variables. Columns 3 and 4 are not pivot columns, which means that z and w are free variables.

We can write the solution to this system as

$$\begin{aligned} x &= -1 + s - t \\ y &= 2 \\ z &= s \\ w &= t \end{aligned}$$

Here the free variables are written as parameters, and the basic variables are given by linear functions of these parameters. ◻

In general, all solutions can be written in terms of the free variables. In such a description, the free variables can take any values (they become parameters), while the basic variables become simple linear functions of these parameters. Indeed, a basic variable x_i is a linear function of *only* those free variables x_j with $j > i$. This leads to the following observation.

Proposition 1.25: Basic and Free Variables

If x_i is a basic variable of a homogeneous system of linear equations, then any solution of the system with $x_j = 0$ for all those free variables x_j with $j > i$ must also have $x_i = 0$.

Using this proposition, we prove a lemma which will be used in the proof of the main result of this section below.

Lemma 1.26: Solutions and the Reduced Row-Echelon Form of a Matrix

Let A and B be two distinct augmented matrices for two homogeneous systems of m equations in n variables, such that A and B are each in reduced row-echelon form. Then, the two systems do not have exactly the same solutions.

Proof. With respect to the linear systems associated with the matrices A and B, there are two cases to consider:

- Case 1: the two systems have the same basic variables
- Case 2: the two systems do not have the same basic variables

In case 1, the two matrices will have exactly the same pivot positions. However, since A and B are not identical, there is some row of A which is different from the corresponding row of B and yet the rows each have a pivot in the same column position. Let i be the index of this column position. Since the matrices are in reduced row-echelon form, the two rows must differ at some entry in a column $j > i$. Let these entries be a in A and b in B, where $a \neq b$. Since A is in reduced row-echelon form, if x_j were a basic variable for its linear system, we would have $a = 0$. Similarly, if x_j were a basic variable for the linear system of the matrix B, we would have $b = 0$. Since a and b are unequal, they cannot both be equal to 0, and hence x_j cannot be a basic variable for both linear systems. However, since the systems have the same basic variables, x_j must then be a free variable for each system. We now look at the solutions of the systems in which x_j is set equal to 1 and all other free variables are set equal to 0. For this choice of parameters, the solution of the system for matrix A has $x_j = -a$, while the solution of the system for matrix B has $x_j = -b$, so that the two systems have different solutions.

In case 2, there is a variable x_i which is a basic variable for one matrix, let's say A, and a free variable for the other matrix B. The system for matrix B has a solution in which $x_i = 1$ and $x_j = 0$ for all other free variables x_j. However, by Proposition 1.25 this cannot be a solution of the system for the matrix A. This completes the proof of case 2. □

Now, we say that the matrix B is **equivalent** to the matrix A provided that B can be obtained from A by performing a sequence of elementary row operations beginning with A. The importance of this concept lies in the following result.

Theorem 1.27: Equivalent Matrices

The two linear systems of equations corresponding to two equivalent augmented matrices have exactly the same solutions.

The proof of this theorem is left as an exercise.

Now, we can use Lemma 1.26 and Theorem 1.27 to prove the main result of this section.

Theorem 1.28: Uniqueness of the Reduced Row-Echelon Form

Every matrix A is equivalent to a unique matrix in reduced row-echelon form.

Proof. Let A be an $m \times n$ matrix and let B and C be matrices in reduced row-echelon form, each equivalent to A. It suffices to show that $B = C$.

Let A^+ be the matrix A augmented with a new rightmost column consisting entirely of zeros. Similarly, augment matrices B and C each with a rightmost column of zeros to obtain B^+ and C^+. Note that B^+ and C^+ are matrices in reduced row-echelon form which are obtained from A^+ by respectively applying the same sequence of elementary row operations which were used to obtain B and C from A.

Now, A^+, B^+, and C^+ can all be considered as augmented matrices of homogeneous linear systems in the variables $x_1, x_2, \cdots, x_n$. Because B^+ and C^+ are each equivalent to A^+, Theorem 1.27 ensures that all three homogeneous linear systems have exactly the same solutions. By Lemma 1.26 we conclude that $B^+ = C^+$. By construction, we must also have $B = C$. □

According to this theorem we can say that each matrix A has a unique reduced row-echelon form.

1.2.4. Rank and Homogeneous Systems

There is a special type of system which requires additional study. This type of system is called a homogeneous system of equations, which we defined above in Definition 1.3. Our focus in this section is to consider what types of solutions are possible for a homogeneous system of equations.

Consider the following definition.

Definition 1.29: Trivial Solution

Consider the homogeneous system of equations given by

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

Then, $x_1 = 0, x_2 = 0, \cdots, x_n = 0$ is always a solution to this system. We call this the ***trivial solution*** *.*

If the system has a solution in which not all of the $x_1, \cdots, x_n$ are equal to zero, then we call this solution **nontrivial** . The trivial solution does not tell us much about the system, as it says that $0 = 0$! Therefore, when working with homogeneous systems of equations, we want to know when the system has a nontrivial solution.

Suppose we have a homogeneous system of m equations, using n variables, and suppose that $n > m$. In other words, there are more variables than equations. Then, it turns out that this system always has a nontrivial solution. Not only will the system have a nontrivial solution, but it also will have infinitely many solutions. It is also possible, but not required, to have a nontrivial solution if $n = m$ and $n < m$.

Consider the following example.

Example 1.30: Solutions to a Homogeneous System of Equations

Find the nontrivial solutions to the following homogeneous system of equations

$$\begin{array}{c}
2x + y - z = 0 \\
x + 2y - 2z = 0
\end{array}$$

Solution. Notice that this system has $m = 2$ equations and $n = 3$ variables, so $n > m$. Therefore by our previous discussion, we expect this system to have infinitely many solutions.

The process we use to find the solutions for a homogeneous system of equations is the same process we used in the previous section. First, we construct the augmented matrix, given by

$$\left[\begin{array}{rrr|r} 2 & 1 & -1 & 0 \\ 1 & 2 & -2 & 0 \end{array}\right]$$

Then, we carry this matrix to its reduced row-echelon form, given below.

$$\left[\begin{array}{rrr|r} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \end{array}\right]$$

The corresponding system of equations is

$$\begin{aligned} x &= 0 \\ y - z &= 0 \end{aligned}$$

Since z is not restrained by any equation, we know that this variable will become our parameter. Let $z = t$ where t is any number. Therefore, our solution has the form

$$\begin{aligned} x &= 0 \\ y &= z = t \\ z &= t \end{aligned}$$

Hence this system has infinitely many solutions, with one parameter t. $\square$

Suppose we were to write the solution to the previous example in another form. Specifically,

$$\begin{aligned} x &= 0 \\ y &= 0 + t \\ z &= 0 + t \end{aligned}$$

can be written as

$$\left[\begin{array}{c} x \\ y \\ z \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \end{array}\right] + t \left[\begin{array}{c} 0 \\ 1 \\ 1 \end{array}\right]$$

Notice that we have constructed a column from the constants in the solution (all equal to 0), as well as a column corresponding to the coefficients on t in each equation. While we will discuss this form of solution more in further chapters, for now consider the column of coefficients of the parameter t. In this case, this is the column $\left[\begin{array}{c} 0 \\ 1 \\ 1 \end{array}\right]$.

There is a special name for this column, which is **basic solution**. The basic solutions of a system are columns constructed from the coefficients on parameters in the solution. We often denote basic solutions by X_1, X_2 etc., depending on how many solutions occur. Therefore, Example 1.30 has the basic solution $X_1 = \left[\begin{array}{c} 0 \\ 1 \\ 1 \end{array}\right]$.

We explore this further in the following example.

Example 1.31: Basic Solutions of a Homogeneous System

Consider the following homogeneous system of equations.

$$\begin{aligned} x + 4y + 3z &= 0 \\ 3x + 12y + 9z &= 0 \end{aligned}$$

Find the basic solutions to this system.

Solution. The augmented matrix of this system and the resulting reduced row-echelon form are

$$\left[\begin{array}{rrr|r} 1 & 4 & 3 & 0 \\ 3 & 12 & 9 & 0 \end{array}\right] \rightarrow \cdots \rightarrow \left[\begin{array}{rrr|r} 1 & 4 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

When written in equations, this system is given by

$$x + 4y + 3z = 0$$

Notice that only x corresponds to a pivot column. In this case, we will have two parameters, one for y and one for z. Let $y = s$ and $z = t$ for any numbers s and t. Then, our solution becomes

$$\begin{aligned} x &= -4s - 3t \\ y &= s \\ z &= t \end{aligned}$$

which can be written as

$$\left[\begin{array}{c} x \\ y \\ z \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \end{array}\right] + s \left[\begin{array}{r} -4 \\ 1 \\ 0 \end{array}\right] + t \left[\begin{array}{r} -3 \\ 0 \\ 1 \end{array}\right]$$

You can see here that we have two columns of coefficients corresponding to parameters, specifically one for s and one for t. Therefore, this system has two basic solutions! These are

$$X_1 = \left[\begin{array}{r} -4 \\ 1 \\ 0 \end{array}\right], X_2 = \left[\begin{array}{r} -3 \\ 0 \\ 1 \end{array}\right]$$

□

We now present a new definition.

Definition 1.32: Linear Combination

*Let $X_1, \cdots, X_n, V$ be column matrices. Then V is said to be a **linear combination** of the columns $X_1, \cdots, X_n$ if there exist scalars, $a_1, \cdots, a_n$ such that*

$$V = a_1 X_1 + \cdots + a_n X_n$$

A remarkable result of this section is that a linear combination of the basic solutions is again a solution to the system. Even more remarkable is that every solution can be written as a linear combination of these solutions. Therefore, if we take a linear combination of the two solutions to Example 1.31, this would also be a solution. For example, we could take the following linear combination

$$3 \left[\begin{array}{r} -4 \\ 1 \\ 0 \end{array}\right] + 2 \left[\begin{array}{r} -3 \\ 0 \\ 1 \end{array}\right] = \left[\begin{array}{r} -18 \\ 3 \\ 2 \end{array}\right]$$

You should take a moment to verify that

$$\left[\begin{array}{c} x \\ y \\ z \end{array}\right] = \left[\begin{array}{r} -18 \\ 3 \\ 2 \end{array}\right]$$

is in fact a solution to the system in Example 1.31.

Another way in which we can find out more information about the solutions of a homogeneous system is to consider the **rank** of the associated coefficient matrix. We now define what is meant by the rank of a matrix.

Definition 1.33: Rank of a Matrix

Let A be a matrix and consider any row-echelon form of A. Then, the number r of leading entries of A does not depend on the row-echelon form you choose, and is called the ***rank*** *of A. We denote it by* $\text{rank}(A)$.

Similarly, we could count the number of pivot positions (or pivot columns) to determine the rank of A.

Example 1.34: Finding the Rank of a Matrix

Consider the matrix

$$\left[\begin{array}{ccc} 1 & 2 & 3 \\ 1 & 5 & 9 \\ 2 & 4 & 6 \end{array}\right]$$

What is its rank?

Solution. First, we need to find the reduced row-echelon form of A. Through the usual algorithm, we find that this is

$$\left[\begin{array}{rrr} \boxed{1} & 0 & -1 \\ 0 & \boxed{1} & 2 \\ 0 & 0 & 0 \end{array}\right]$$

Here we have two leading entries, or two pivot positions, shown above in boxes.The rank of A is $r = 2$. ∎

Notice that we would have achieved the same answer if we had found the row-echelon form of A instead of the reduced row-echelon form.

Suppose we have a homogeneous system of m equations in n variables, and suppose that $n > m$. From our above discussion, we know that this system will have infinitely many solutions. If we consider the rank of the coefficient matrix of this system, we can find out even more about the solution. Note that we are looking at just the coefficient matrix, not the entire augmented matrix.

Theorem 1.35: Rank and Solutions to a Homogeneous System

Let A be the $m \times n$ coefficient matrix corresponding to a homogeneous system of equations, and suppose A has rank r. Then, the solution to the corresponding system has $n - r$ parameters.

Consider our above Example 1.31 in the context of this theorem. The system in this example has $m = 2$ equations in $n = 3$ variables. First, because $n > m$, we know that the system has a nontrivial solution, and therefore infinitely many solutions. This tells us that the solution will contain at least one parameter. The rank of the coefficient matrix can tell us even more about the solution! The rank of the coefficient matrix of the system is 1, as it has one leading entry in row-echelon form. Theorem 1.35 tells us that the solution will have $n - r = 3 - 1 = 2$ parameters. You can check that this is true in the solution to Example 1.31.

Notice that if $n = m$ or $n < m$, it is possible to have either a unique solution (which will be the trivial solution) or infinitely many solutions.

We are not limited to homogeneous systems of equations here. The rank of a matrix can be used to learn about the solutions of any system of linear equations. In the previous section, we discussed that a system of equations can have no solution, a unique solution, or infinitely many solutions. Suppose the system is consistent, whether it is homogeneous or not. The following theorem tells us how we can use the rank to learn about the type of solution we have.

Theorem 1.36: Rank and Solutions to a Consistent System of Equations

Let A be the $m \times (n + 1)$ augmented matrix corresponding to a consistent system of equations in n variables, and suppose A has rank r. Then

1. *the system has a unique solution if $r = n$*
2. *the system has infinitely many solutions if $r < n$*

We will not present a formal proof of this, but consider the following discussions.

1. *No Solution* The above theorem assumes that the system is consistent, that is, that it has a solution. It turns out that it is possible for the augmented matrix of a system with no solution to have any rank r as long as $r > 1$. Therefore, we must know that the system is consistent in order to use this theorem!

2. *Unique Solution* Suppose $r = n$. Then, there is a pivot position in every column of the coefficient matrix of A. Hence, there is a unique solution.

3. *Infinitely Many Solutions* Suppose $r < n$. Then there are infinitely many solutions. There are less pivot positions (and hence less leading entries) than columns, meaning that not every column is a pivot column. The columns which are *not* pivot columns correspond to parameters. In fact, in this case we have $n - r$ parameters.

Exercises

Exercise 1.2.1 *Find the point* (x_1, y_1) *which lies on both lines,* $x + 3y = 1$ *and* $4x - y = 3$.

Exercise 1.2.2 *Find the point of intersection of the two lines* $3x + y = 3$ *and* $x + 2y = 1$.

Exercise 1.2.3 *Do the three lines,* $x + 2y = 1, 2x - y = 1,$ *and* $4x + 3y = 3$ *have a common point of intersection? If so, find the point and if not, tell why they don't have such a common point of intersection.*

Exercise 1.2.4 *Do the three planes,* $x + y - 3z = 2$, $2x + y + z = 1$, *and* $3x + 2y - 2z = 0$ *have a common point of intersection? If so, find one and if not, tell why there is no such point.*

Exercise 1.2.5 *Four times the weight of Gaston is 150 pounds more than the weight of Ichabod. Four times the weight of Ichabod is 660 pounds less than seventeen times the weight of Gaston. Four times the weight of Gaston plus the weight of Siegfried equals 290 pounds. Brunhilde would balance all three of the others. Find the weights of the four people.*

Exercise 1.2.6 *Consider the following augmented matrix in which* $*$ *denotes an arbitrary number and* $\blacksquare$ *denotes a nonzero number. Determine whether the given augmented matrix is consistent. If consistent, is the solution unique?*

$$\left[\begin{array}{ccccc|c} \blacksquare & * & * & * & * & * \\ 0 & \blacksquare & * & * & 0 & * \\ 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * \end{array}\right]$$

Exercise 1.2.7 *Consider the following augmented matrix in which* $*$ *denotes an arbitrary number and* $\blacksquare$ *denotes a nonzero number. Determine whether the given augmented matrix is consistent. If consistent, is the solution unique?*

$$\left[\begin{array}{ccc|c} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & \blacksquare & * \end{array}\right]$$

Exercise 1.2.8 *Consider the following augmented matrix in which* $*$ *denotes an arbitrary number and* $\blacksquare$ *denotes a nonzero number. Determine whether the given augmented matrix is consistent. If consistent, is the solution unique?*

$$\left[\begin{array}{ccccc|c} \blacksquare & * & * & * & * & * \\ 0 & \blacksquare & 0 & * & 0 & * \\ 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * \end{array}\right]$$

Exercise 1.2.9 *Consider the following augmented matrix in which $*$ denotes an arbitrary number and $\blacksquare$ denotes a nonzero number. Determine whether the given augmented matrix is consistent. If consistent, is the solution unique?*

$$\left[\begin{array}{ccccc|c} \blacksquare & * & * & * & * & * \\ 0 & \blacksquare & * & * & 0 & * \\ 0 & 0 & 0 & 0 & \blacksquare & 0 \\ 0 & 0 & 0 & 0 & * & \blacksquare \end{array}\right]$$

Exercise 1.2.10 *Suppose a system of equations has fewer equations than variables. Will such a system necessarily be consistent? If so, explain why and if not, give an example which is not consistent.*

Exercise 1.2.11 *If a system of equations has more equations than variables, can it have a solution? If so, give an example and if not, tell why not.*

Exercise 1.2.12 *Find h such that*

$$\left[\begin{array}{cc|c} 2 & h & 4 \\ 3 & 6 & 7 \end{array}\right]$$

is the augmented matrix of an inconsistent system.

Exercise 1.2.13 *Find h such that*

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

is the augmented matrix of a consistent system.

Exercise 1.2.14 *Find h such that*

$$\left[\begin{array}{cc|c} 1 & 1 & 4 \\ 3 & h & 12 \end{array}\right]$$

is the augmented matrix of a consistent system.

Exercise 1.2.15 *Choose h and k such that the augmented matrix shown has each of the following:*

1. *one solution*
2. *no solution*
3. *infinitely many solutions*

$$\left[\begin{array}{cc|c} 1 & h & 2 \\ 2 & 4 & k \end{array}\right]$$

Exercise 1.2.16 *Choose h and k such that the augmented matrix shown has each of the following:*

1. *one solution*

2. *no solution*

3. *infinitely many solutions*

$$\left[\begin{array}{cc|c} 1 & 2 & 2 \\ 2 & h & k \end{array}\right]$$

Exercise 1.2.17 *Determine if the system is consistent. If so, is the solution unique?*

$$\begin{gathered} x+2y+z-w=2 \\ x-y+z+w=1 \\ 2x+y-z=1 \\ 4x+2y+z=5 \end{gathered}$$

Exercise 1.2.18 *Determine if the system is consistent. If so, is the solution unique?*

$$\begin{gathered} x+2y+z-w=2 \\ x-y+z+w=0 \\ 2x+y-z=1 \\ 4x+2y+z=3 \end{gathered}$$

Exercise 1.2.19 *Determine which matrices are in reduced row-echelon form.*

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

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

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

Exercise 1.2.20 *Row reduce the following matrix to obtain the row-echelon form. Then continue to obtain the reduced row-echelon form.*

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

Exercise 1.2.21 *Row reduce the following matrix to obtain the row-echelon form. Then continue to obtain the reduced row-echelon form.*

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

Exercise 1.2.22 *Row reduce the following matrix to obtain the row-echelon form. Then continue to obtain the reduced row-echelon form.*

$$\left[\begin{array}{rrrr} 3 & -6 & -7 & -8 \\ 1 & -2 & -2 & -2 \\ 1 & -2 & -3 & -4 \end{array}\right]$$

Exercise 1.2.23 *Row reduce the following matrix to obtain the row-echelon form. Then continue to obtain the reduced row-echelon form.*

$$\left[\begin{array}{rrrr} 2 & 4 & 5 & 15 \\ 1 & 2 & 3 & 9 \\ 1 & 2 & 2 & 6 \end{array}\right]$$

Exercise 1.2.24 *Row reduce the following matrix to obtain the row-echelon form. Then continue to obtain the reduced row-echelon form.*

$$\left[\begin{array}{rrrr} 4 & -1 & 7 & 10 \\ 1 & 0 & 3 & 3 \\ 1 & -1 & -2 & 1 \end{array}\right]$$

Exercise 1.2.25 *Row reduce the following matrix to obtain the row-echelon form. Then continue to obtain the reduced row-echelon form.*

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

Exercise 1.2.26 *Row reduce the following matrix to obtain the row-echelon form. Then continue to obtain the reduced row-echelon form.*

$$\left[\begin{array}{rrrr} -2 & 3 & -8 & 7 \\ 1 & -2 & 5 & -5 \\ 1 & -3 & 7 & -8 \end{array}\right]$$

Exercise 1.2.27 *Find the solution of the system whose augmented matrix is*

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

Exercise 1.2.28 *Find the solution of the system whose augmented matrix is*

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

Exercise 1.2.29 *Find the solution of the system whose augmented matrix is*

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

Exercise 1.2.30 *Find the solution of the system whose augmented matrix is*

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

Exercise 1.2.31 *Find the solution of the system whose augmented matrix is*

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

Exercise 1.2.32 *Find the solution to the system of equations,* $7x + 14y + 15z = 22$, $2x + 4y + 3z = 5$, *and* $3x + 6y + 10z = 13$.

Exercise 1.2.33 *Find the solution to the system of equations,* $3x - y + 4z = 6$, $y + 8z = 0$, *and* $-2x + y = -4$.

Exercise 1.2.34 *Find the solution to the system of equations,* $9x - 2y + 4z = -17$, $13x - 3y + 6z = -25$, *and* $-2x - z = 3$.

Exercise 1.2.35 *Find the solution to the system of equations,* $65x + 84y + 16z = 546$, $81x + 105y + 20z = 682$, *and* $84x + 110y + 21z = 713$.

Exercise 1.2.36 *Find the solution to the system of equations,* $8x+2y+3z = -3, 8x+3y+3z = -1$, *and* $4x + y + 3z = -9$.

Exercise 1.2.37 *Find the solution to the system of equations,* $-8x + 2y + 5z = 18, -8x + 3y + 5z = 13$, *and* $-4x + y + 5z = 19$.

Exercise 1.2.38 *Find the solution to the system of equations,* $3x - y - 2z = 3$, $y - 4z = 0$, *and* $-2x + y = -2$.

Exercise 1.2.39 *Find the solution to the system of equations,* $-9x+15y = 66, -11x+18y = 79$, $-x + y = 4$, *and* $z = 3$.

Exercise 1.2.40 *Find the solution to the system of equations,* $-19x + 8y = -108$, $-71x + 30y = -404$, $-2x + y = -12$, $4x + z = 14$.

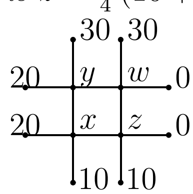
Exercise 1.2.41 *Suppose a system of equations has fewer equations than variables and you have found a solution to this system of equations. Is it possible that your solution is the only one? Explain.*

Exercise 1.2.42 *Suppose a system of linear equations has a* 2×4 *augmented matrix and the last column is a pivot column. Could the system of linear equations be consistent? Explain.*

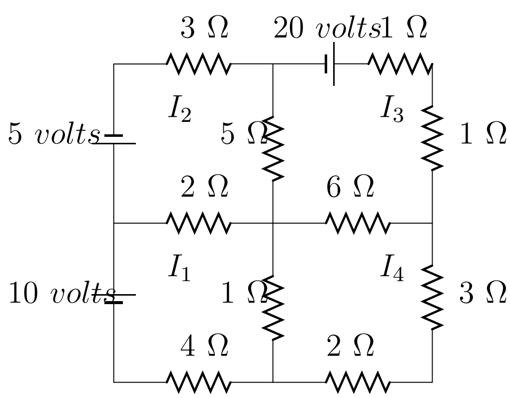
Exercise 1.2.43 *Suppose the coefficient matrix of a system of* n *equations with* n *variables has the property that every column is a pivot column. Does it follow that the system of equations must have a solution? If so, must the solution be unique? Explain.*

Exercise 1.2.44 *Suppose there is a unique solution to a system of linear equations. What must be true of the pivot columns in the augmented matrix?*

Exercise 1.2.45 *The steady state temperature,* u, *of a plate solves Laplace's equation,* $\Delta u = 0$. *One way to approximate the solution is to divide the plate into a square mesh and require the temperature at each node to equal the average of the temperature at the four adjacent nodes. In the following picture, the numbers represent the observed temperature at the indicated nodes. Find the temperature at the interior nodes, indicated by* x, y, z, *and* w. *One of the equations is* $z = \frac{1}{4}(10 + 0 + w + x)$.



Exercise 1.2.46 *Consider the following diagram of four circuits.*



The jagged lines denote resistors and the numbers next to them give their resistance in ohms, written as Ω. *The breaks in the lines having one short line and one long line denote a voltage source which causes the current to flow in the direction which goes from the longer of the two lines toward the shorter along the unbroken part of the circuit. The current in amps in the four circuits is denoted by* I_1, I_2, I_3, I_4 *and it is understood that the motion is in* *the counter clockwise direction. If I_k ends up being negative, then it just means the current flows in the clockwise direction. Then Kirchhoff's law states:*

The sum of the resistance times the amps in the counter clockwise direction around a loop equals the sum of the voltage sources in the same direction around the loop.

In the above diagram, the top left circuit should give the equation

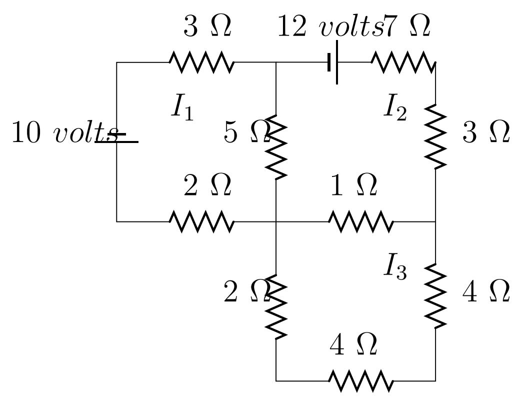
$$2I_2 - 2I_1 + 5I_2 - 5I_3 + 3I_2 = 5$$

For the circuit on the lower left, you should have

$$4I_1 + I_1 - I_4 + 2I_1 - 2I_2 = -10$$

Write equations for each of the other two circuits and then give a solution to the resulting system of equations.

Exercise 1.2.47 *Consider the following diagram of three circuits.*



The jagged lines denote resistors and the numbers next to them give their resistance in ohms, written as Ω. The breaks in the lines having one short line and one long line denote a voltage source which causes the current to flow in the direction which goes from the longer of the two lines toward the shorter along the unbroken part of the circuit. The current in amps in the four circuits is denoted by I_1, I_2, I_3 and it is understood that the motion is in the counter clockwise direction. If I_k ends up being negative, then it just means the current flows in the clockwise direction. Then Kirchhoff's law states:

The sum of the resistance times the amps in the counter clockwise direction around a loop equals the sum of the voltage sources in the same direction around the loop.

Find I_1, I_2, I_3.

Exercise 1.2.48 *Find the rank of the following matrix.*

$$\begin{bmatrix} 4 & -16 & -1 & -5 \\ 1 & -4 & 0 & -1 \\ 1 & -4 & -1 & -2 \end{bmatrix}$$

Exercise 1.2.49 *Find the rank of the following matrix.*

$$\begin{bmatrix} 3 & 6 & 5 & 12 \\ 1 & 2 & 2 & 5 \\ 1 & 2 & 1 & 2 \end{bmatrix}$$

Exercise 1.2.50 *Find the rank of the following matrix.*

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

Exercise 1.2.51 *Find the rank of the following matrix.*

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

Exercise 1.2.52 *Find the rank of the following matrix.*

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

Exercise 1.2.53 *Find the rank of the following matrix.*

$$\begin{bmatrix} 4 & 15 & 29 \\ 1 & 4 & 8 \\ 1 & 3 & 5 \\ 3 & 9 & 15 \end{bmatrix}$$

Exercise 1.2.54 *Find the rank of the following matrix.*

$$\begin{bmatrix} 0 & 0 & -1 & 0 & 1 \\ 1 & 2 & 3 & -2 & -18 \\ 1 & 2 & 2 & -1 & -11 \\ -1 & -2 & -2 & 1 & 11 \end{bmatrix}$$

Exercise 1.2.55 *Find the rank of the following matrix.*

$$\begin{bmatrix} 1 & -2 & 0 & 3 & 11 \\ 1 & -2 & 0 & 4 & 15 \\ 1 & -2 & 0 & 3 & 11 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Exercise 1.2.56 *Find the rank of the following matrix.*

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

Exercise 1.2.57 *Find the rank of the following matrix.*

$$\begin{bmatrix} 4 & 4 & 20 & -1 & 17 \\ 1 & 1 & 5 & 0 & 5 \\ 1 & 1 & 5 & -1 & 2 \\ 3 & 3 & 15 & -3 & 6 \end{bmatrix}$$

Exercise 1.2.58 *Find the rank of the following matrix.*

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

Exercise 1.2.59 *Suppose A is an $m \times n$ matrix. Explain why the rank of A is always no larger than* $\min(m, n)$.

Exercise 1.2.60 *State whether each of the following sets of data are possible for the matrix equation $AX = B$. If possible, describe the solution set. That is, tell whether there exists a unique solution, no solution or infinitely many solutions. Here, $[A|B]$ denotes the augmented matrix.*

1. *A is a 5×6 matrix, rank$(A) = 4$ and rank$[A|B] = 4$.*

2. *A is a 3×4 matrix, rank$(A) = 3$ and rank$[A|B] = 2$.*

3. *A is a 4×2 matrix, rank$(A) = 4$ and rank$[A|B] = 4$.*

4. *A is a 5×5 matrix, rank$(A) = 4$ and rank$[A|B] = 5$.*

5. *A is a 4×2 matrix, rank$(A) = 2$ and rank$[A|B] = 2$.*

Exercise 1.2.61 *Consider the system $-5x+2y-z = 0$ and $-5x-2y-z = 0$. Both equations equal zero and so $-5x + 2y - z = -5x - 2y - z$ which is equivalent to $y = 0$. Does it follow that x and z can equal anything? Notice that when $x = 1$, $z = -4$, and $y = 0$ are plugged in to the equations, the equations do not equal 0. Why?*

2. Matrices

2.1 Matrix Arithmetic

Outcomes

A. *Perform the matrix operations of matrix addition, scalar multiplication, transposition and matrix multiplication. Identify when these operations are not defined. Represent these operations in terms of the entries of a matrix.*

B. *Prove algebraic properties for matrix addition, scalar multiplication, transposition, and matrix multiplication. Apply these properties to manipulate an algebraic expression involving matrices.*

C. *Compute the inverse of a matrix using row operations, and prove identities involving matrix inverses.*

E. *Solve a linear system using matrix algebra.*

F. *Use multiplication by an elementary matrix to apply row operations.*

G. *Write a matrix as a product of elementary matrices.*

You have now solved systems of equations by writing them in terms of an augmented matrix and then doing row operations on this augmented matrix. It turns out that matrices are important not only for systems of equations but also in many applications.

Recall that a **matrix** is a rectangular array of numbers. Several of them are referred to as **matrices**. For example, here is a matrix.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{bmatrix} \tag{2.1}$$

Recall that the size or dimension of a matrix is defined as $m \times n$ where m is the number of rows and n is the number of columns. The above matrix is a 3×4 matrix because there are three rows and four columns. You can remember the columns are like columns in a Greek temple. They stand upright while the rows lay flat like rows made by a tractor in a plowed field.

When specifying the size of a matrix, you always list the number of rows before the number of columns.You might remember that you always list the rows before the columns by using the phrase **Row**man **C**atholic.

Consider the following definition.

Definition 2.1: Square Matrix

A matrix A which has size $n \times n$ is called a ***square matrix*** *. In other words, A is a square matrix if it has the same number of rows and columns.*

There is some notation specific to matrices which we now introduce. We denote the columns of a matrix A by A_j as follows

$$A = \left[\begin{array}{cccc} A_1 & A_2 & \cdots & A_n \end{array} \right]$$

Therefore, A_j is the j^{th} column of A, when counted from left to right.

The individual elements of the matrix are called **entries** or **components** of A. Elements of the matrix are identified according to their position. The $(\mathbf{i}, \mathbf{j})$**-entry** of a matrix is the entry in the i^{th} row and j^{th} column. For example, in the matrix 2.1 above, 8 is in position $(2,3)$ (and is called the $(2,3)$-entry) because it is in the second row and the third column.

In order to remember which matrix we are speaking of, we will denote the entry in the i^{th} row and the j^{th} column of matrix A by a_{ij}. Then, we can write A in terms of its entries, as $A = [a_{ij}]$. Using this notation on the matrix in 2.1, $a_{23} = 8, a_{32} = -9, a_{12} = 2$, etc.

There are various operations which are done on matrices of appropriate sizes. Matrices can be added to and subtracted from other matrices, multiplied by a scalar, and multiplied by other matrices. We will never divide a matrix by another matrix, but we will see later how matrix inverses play a similar role.

In doing arithmetic with matrices, we often define the action by what happens in terms of the entries (or components) of the matrices. Before looking at these operations in depth, consider a few general definitions.

Definition 2.2: The Zero Matrix

The $m \times n$ ***zero matrix*** *is the $m \times n$ matrix having every entry equal to zero. It is denoted by* 0.

One possible zero matrix is shown in the following example.

Example 2.3: The Zero Matrix

The 2×3 zero matrix is $0 = \left[\begin{array}{ccc} 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right]$.

Note there is a 2×3 zero matrix, a 3×4 zero matrix, etc. In fact there is a zero matrix for every size!

Definition 2.4: Equality of Matrices

Let A and B be two $m \times n$ matrices. Then $A = B$ means that for $A = [a_{ij}]$ and $B = [b_{ij}]$, $a_{ij} = b_{ij}$ for all $1 \leq i \leq m$ and $1 \leq j \leq n$.

In other words, two matrices are equal exactly when they are the same size and the corresponding entries are identical. Thus

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

because they are different sizes. Also,

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

because, although they are the same size, their corresponding entries are not identical.

In the following section, we explore addition of matrices.

2.1.1. Addition of Matrices

When adding matrices, all matrices in the sum need have the same size. For example,

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

and

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

cannot be added, as one has size 3×2 while the other has size 2×3.

However, the addition

$$\begin{bmatrix} 4 & 6 & 3 \\ 5 & 0 & 4 \\ 11 & -2 & 3 \end{bmatrix} + \begin{bmatrix} 0 & 5 & 0 \\ 4 & -4 & 14 \\ 1 & 2 & 6 \end{bmatrix}$$

is possible.

The formal definition is as follows.

Definition 2.5: Addition of Matrices

Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two $m \times n$ matrices. Then $A + B = C$ where C is the $m \times n$ matrix $C = [c_{ij}]$ defined by

$$c_{ij} = a_{ij} + b_{ij}$$

This definition tells us that when adding matrices, we simply add corresponding entries of the matrices. This is demonstrated in the next example.

Example 2.6: Addition of Matrices of Same Size

Add the following matrices, if possible.

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

Solution. Notice that both A and B are of size 2×3. Since A and B are of the same size, the addition is possible. Using Definition 2.5, the addition is done as follows.

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

♠

Addition of matrices obeys very much the same properties as normal addition with numbers. Note that when we write for example $A + B$ then we assume that both matrices are of equal size so that the operation is indeed possible.

Proposition 2.7: Properties of Matrix Addition

Let A, B and C be matrices. Then, the following properties hold.

- *Commutative Law of Addition*

$$A + B = B + A \tag{2.2}$$

- *Associative Law of Addition*

$$(A + B) + C = A + (B + C) \tag{2.3}$$

- *Existence of an Additive Identity*

$$\begin{array}{c} \text{There exists a zero matrix 0 such that} \\ A + 0 = A \end{array} \tag{2.4}$$

- *Existence of an Additive Inverse*

$$\begin{array}{c} \text{There exists a matrix } -A \text{ such that} \\ A + (-A) = 0 \end{array} \tag{2.5}$$

Proof. Consider the Commutative Law of Addition given in 2.2. Let A, B, C, and D be matrices such that $A + B = C$ and $B + A = D$. We want to show that $D = C$. To do so, we will use the definition of matrix addition given in Definition 2.5. Now,

$$c_{ij} = a_{ij} + b_{ij} = b_{ij} + a_{ij} = d_{ij}$$

Therefore, $C = D$ because the ij^{th} entries are the same for all i and j. Note that the conclusion follows from the commutative law of addition of numbers, which says that if a and b are two numbers, then $a + b = b + a$. The proof of the other results are similar, and are left as an exercise. ▯

We call the zero matrix in 2.4 the **additive identity**. Similarly, we call the matrix $-A$ in 2.5 the **additive inverse**. $-A$ is defined to equal $(-1)\,A = [-a_{ij}]$. In other words, every entry of A is multiplied by -1. In the next section we will study scalar multiplication in more depth to understand what is meant by $(-1)\,A$.

2.1.2. Scalar Multiplication of Matrices

Recall that we use the word *scalar* when referring to numbers. Therefore, *scalar multiplication of a matrix* is the multiplication of a matrix by a number. To illustrate this concept, consider the following example in which a matrix is multiplied by the scalar 3.

$$3\left[\begin{array}{rrrr} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{array}\right] = \left[\begin{array}{rrrr} 3 & 6 & 9 & 12 \\ 15 & 6 & 24 & 21 \\ 18 & -27 & 3 & 6 \end{array}\right]$$

The new matrix is obtained by multiplying every entry of the original matrix by the given scalar.

The formal definition of scalar multiplication is as follows.

Definition 2.8: Scalar Multiplication of Matrices

If $A = [a_{ij}]$ and k is a scalar, then $kA = [ka_{ij}]$.

Consider the following example.

Example 2.9: Effect of Multiplication by a Scalar

Find the result of multiplying the following matrix A by 7.

$$A = \left[\begin{array}{rr} 2 & 0 \\ 1 & -4 \end{array}\right]$$

Solution. By Definition 2.8, we multiply each element of A by 7. Therefore,

$$7A = 7\left[\begin{array}{rr} 2 & 0 \\ 1 & -4 \end{array}\right] = \left[\begin{array}{rr} 7(2) & 7(0) \\ 7(1) & 7(-4) \end{array}\right] = \left[\begin{array}{rr} 14 & 0 \\ 7 & -28 \end{array}\right]$$

▯

Similarly to addition of matrices, there are several properties of scalar multiplication which hold.

Proposition 2.10: Properties of Scalar Multiplication

Let A, B be matrices, and k, p be scalars. Then, the following properties hold.

- *Distributive Law over Matrix Addition*

$$k(A+B) = kA + kB$$

- *Distributive Law over Scalar Addition*

$$(k+p)A = kA + pA$$

- *Associative Law for Scalar Multiplication*

$$k(pA) = (kp)A$$

- *Rule for Multiplication by 1*

$$1A = A$$

The proof of this proposition is similar to the proof of Proposition 2.7 and is left an exercise to the reader.

2.1.3. Multiplication of Matrices

The next important matrix operation we will explore is multiplication of matrices. The operation of matrix multiplication is one of the most important and useful of the matrix operations. Throughout this section, we will also demonstrate how matrix multiplication relates to linear systems of equations.

First, we provide a formal definition of row and column vectors.

Definition 2.11: Row and Column Vectors

Matrices of size $n \times 1$ or $1 \times n$ are called ***vectors****. If X is such a matrix, then we write x_i to denote the entry of X in the i^{th} row of a column matrix, or the i^{th} column of a row matrix.*
The $n \times 1$ matrix

$$X = \left[\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array} \right]$$

is called a ***column vector.*** *The $1 \times n$ matrix*

$$X = \left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array} \right]$$

is called a ***row vector****.*

We may simply use the term **vector** throughout this text to refer to either a column or row vector. If we do so, the context will make it clear which we are referring to.

In this chapter, we will again use the notion of linear combination of vectors as in Definition 4.7. In this context, a linear combination is a sum consisting of vectors multiplied by scalars. For example,

$$\left[\begin{array}{r} 50 \\ 122 \end{array}\right] = 7\left[\begin{array}{r} 1 \\ 4 \end{array}\right] + 8\left[\begin{array}{r} 2 \\ 5 \end{array}\right] + 9\left[\begin{array}{r} 3 \\ 6 \end{array}\right]$$

is a linear combination of three vectors.

It turns out that we can express any system of linear equations as a linear combination of vectors. In fact, the vectors that we will use are just the columns of the corresponding augmented matrix!

Definition 2.12: The Vector Form of a System of Linear Equations

Suppose we have a system of equations given by

$$\begin{array}{c} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{array}$$

We can express this system in ***vector form*** *which is as follows:*

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

Notice that each vector used here is one column from the corresponding augmented matrix. There is one vector for each variable in the system, along with the constant vector.

The first important form of matrix multiplication is multiplying a matrix by a vector. Consider the product given by

$$\left[\begin{array}{rrr} 1 & 2 & 3 \\ 4 & 5 & 6 \end{array}\right]\left[\begin{array}{r} 7 \\ 8 \\ 9 \end{array}\right]$$

We will soon see that this equals

$$7\left[\begin{array}{c} 1 \\ 4 \end{array}\right] + 8\left[\begin{array}{c} 2 \\ 5 \end{array}\right] + 9\left[\begin{array}{c} 3 \\ 6 \end{array}\right] = \left[\begin{array}{c} 50 \\ 122 \end{array}\right]$$

In general terms,

$$\begin{aligned} \left[\begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{array}\right]\left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\right] &= x_1\left[\begin{array}{c} a_{11} \\ a_{21} \end{array}\right] + x_2\left[\begin{array}{c} a_{12} \\ a_{22} \end{array}\right] + x_3\left[\begin{array}{c} a_{13} \\ a_{23} \end{array}\right] \\ &= \left[\begin{array}{c} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \end{array}\right] \end{aligned}$$

Thus you take x_1 times the first column, add to x_2 times the second column, and finally x_3 times the third column. The above sum is a linear combination of the columns of the matrix. When you multiply a matrix on the left by a vector on the right, the numbers making up the vector are just the scalars to be used in the linear combination of the columns as illustrated above.

Here is the formal definition of how to multiply an $m \times n$ matrix by an $n \times 1$ column vector.

Definition 2.13: Multiplication of Vector by Matrix

Let $A = [a_{ij}]$ be an $m \times n$ matrix and let X be an $n \times 1$ matrix given by

$$A = [A_1 \cdots A_n], X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Then the product AX is the $m \times 1$ column vector which equals the following linear combination of the columns of A:

$$x_1 A_1 + x_2 A_2 + \cdots + x_n A_n = \sum_{j=1}^{n} x_j A_j$$

If we write the columns of A in terms of their entries, they are of the form

$$A_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

Then, we can write the product AX as

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

Note that multiplication of an $m \times n$ matrix and an $n \times 1$ vector produces an $m \times 1$ vector.

Here is an example.

Example 2.14: A Vector Multiplied by a Matrix

Compute the product AX for

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

Solution. We will use Definition 2.13 to compute the product. Therefore, we compute the product AX as follows.

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

□

Using the above operation, we can also write a system of linear equations in **matrix form**. In this form, we express the system as a matrix multiplied by a vector. Consider the following definition.

Definition 2.15: The Matrix Form of a System of Linear Equations

Suppose we have a system of equations given by

$$\begin{gathered} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{gathered}$$

Then we can express this system in ***matrix form*** *as follows.*

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

The expression $AX = B$ is also known as the **Matrix Form** of the corresponding system of linear equations. The matrix A is simply the coefficient matrix of the system, the vector

X is the column vector constructed from the variables of the system, and finally the vector B is the column vector constructed from the constants of the system. It is important to note that any system of linear equations can be written in this form.

Notice that if we write a homogeneous system of equations in matrix form, it would have the form $AX = 0$, for the zero vector 0.

You can see from this definition that a vector

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

will satisfy the equation $AX = B$ only when the entries $x_1, x_2, \cdots, x_n$ of the vector X are solutions to the original system.

Now that we have examined how to multiply a matrix by a vector, we wish to consider the case where we multiply two matrices of more general sizes, although these sizes still need to be appropriate as we will see. For example, in Example 2.14, we multiplied a 3×4 matrix by a 4×1 vector. We want to investigate how to multiply other sizes of matrices.

We have not yet given any conditions on when matrix multiplication is possible! For matrices A and B, in order to form the product AB, the number of columns of A must equal the number of rows of B. Consider a product AB where A has size $m \times n$ and B has size $n \times p$. Then, the product in terms of size of matrices is given by

$$(m \times \overbrace{n)\;(n}^{\text{these must match!}} \times p\;) = m \times p$$

Note the two outside numbers give the size of the product. One of the most important rules regarding matrix multiplication is the following. If the two middle numbers don't match, you can't multiply the matrices!

When the number of columns of A equals the number of rows of B the two matrices are said to be **conformable** and the product AB is obtained as follows.

Definition 2.16: Multiplication of Two Matrices

Let A be an $m \times n$ matrix and let B be an $n \times p$ matrix of the form

$$B = [B_1 \cdots B_p]$$

where $B_1, ..., B_p$ are the $n \times 1$ columns of B. Then the $m \times p$ matrix AB is defined as follows:

$$AB = A\,[B_1 \cdots B_p] = [(AB)_1 \cdots (AB)_p]$$

where $(AB)_k$ is an $m \times 1$ matrix or column vector which gives the k^{th} column of AB.

Consider the following example.

Example 2.17: Multiplying Two Matrices

Find AB if possible.

$$A = \left[\begin{array}{rrr} 1 & 2 & 1 \\ 0 & 2 & 1 \end{array} \right], B = \left[\begin{array}{rrr} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{array} \right]$$

Solution. The first thing you need to verify when calculating a product is whether the multiplication is possible. The first matrix has size 2×3 and the second matrix has size 3×3. The inside numbers are equal, so A and B are conformable matrices. According to the above discussion AB will be a 2×3 matrix. Definition 2.16 gives us a way to calculate each column of AB, as follows.

$$\left[\overbrace{\left[\begin{array}{rrr} 1 & 2 & 1 \\ 0 & 2 & 1 \end{array} \right] \left[\begin{array}{r} 1 \\ 0 \\ -2 \end{array} \right]}^{\text{First column}}, \overbrace{\left[\begin{array}{rrr} 1 & 2 & 1 \\ 0 & 2 & 1 \end{array} \right] \left[\begin{array}{r} 2 \\ 3 \\ 1 \end{array} \right]}^{\text{Second column}}, \overbrace{\left[\begin{array}{rrr} 1 & 2 & 1 \\ 0 & 2 & 1 \end{array} \right] \left[\begin{array}{r} 0 \\ 1 \\ 1 \end{array} \right]}^{\text{Third column}} \right]$$

You know how to multiply a matrix times a vector, using Definition 2.13 for each of the three columns. Thus

$$\left[\begin{array}{rrr} 1 & 2 & 1 \\ 0 & 2 & 1 \end{array} \right] \left[\begin{array}{rrr} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{array} \right] = \left[\begin{array}{rrr} -1 & 9 & 3 \\ -2 & 7 & 3 \end{array} \right]$$

♢

Since vectors are simply $n \times 1$ or $1 \times m$ matrices, we can also multiply a vector by another vector.

Example 2.18: Vector Times Vector Multiplication

Multiply if possible $\left[\begin{array}{r} 1 \\ 2 \\ 1 \end{array} \right] \left[\begin{array}{rrrr} 1 & 2 & 1 & 0 \end{array} \right]$.

Solution. In this case we are multiplying a matrix of size 3×1 by a matrix of size 1×4. The inside numbers match so the product is defined. Note that the product will be a matrix of size 3×4. Using Definition 2.16, we can compute this product as follows

$$\left[\begin{array}{r} 1 \\ 2 \\ 1 \end{array} \right] \left[\begin{array}{rrrr} 1 & 2 & 1 & 0 \end{array} \right] = \left[\overbrace{\left[\begin{array}{r} 1 \\ 2 \\ 1 \end{array} \right] \left[\begin{array}{r} 1 \end{array} \right]}^{\text{First column}}, \overbrace{\left[\begin{array}{r} 1 \\ 2 \\ 1 \end{array} \right] \left[\begin{array}{r} 2 \end{array} \right]}^{\text{Second column}}, \overbrace{\left[\begin{array}{r} 1 \\ 2 \\ 1 \end{array} \right] \left[\begin{array}{r} 1 \end{array} \right]}^{\text{Third column}}, \overbrace{\left[\begin{array}{r} 1 \\ 2 \\ 1 \end{array} \right] \left[\begin{array}{r} 0 \end{array} \right]}^{\text{Fourth column}} \right]$$

You can use Definition 2.13 to verify that this product is

$$\left[\begin{array}{cccc} 1 & 2 & 1 & 0 \\ 2 & 4 & 2 & 0 \\ 1 & 2 & 1 & 0 \end{array}\right]$$

♢

Example 2.19: A Multiplication Which is Not Defined

Find BA if possible.

$$B=\left[\begin{array}{ccc} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{array}\right], A=\left[\begin{array}{ccc} 1 & 2 & 1 \\ 0 & 2 & 1 \end{array}\right]$$

Solution. First check if it is possible. This product is of the form $(3\times 3)(2\times 3)$. The inside numbers do not match and so you can't do this multiplication. ♢

In this case, we say that the multiplication is not defined. Notice that these are the same matrices which we used in Example 2.17. In this example, we tried to calculate BA instead of AB. This demonstrates another property of matrix multiplication. While the product AB maybe be defined, we cannot assume that the product BA will be possible. Therefore, it is important to always check that the product is defined before carrying out any calculations.

Earlier, we defined the zero matrix 0 to be the matrix (of appropriate size) containing zeros in all entries. Consider the following example for multiplication by the zero matrix.

Example 2.20: Multiplication by the Zero Matrix

Compute the product $A0$ for the matrix

$$A=\left[\begin{array}{cc} 1 & 2 \\ 3 & 4 \end{array}\right]$$

and the 2×2 zero matrix given by

$$0=\left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array}\right]$$

Solution. In this product, we compute

$$\left[\begin{array}{cc} 1 & 2 \\ 3 & 4 \end{array}\right]\left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array}\right]=\left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array}\right]$$

Hence, $A0=0$. ♢

Notice that we could also multiply A by the 2×1 zero vector given by $\left[\begin{array}{c} 0 \\ 0 \end{array}\right]$. The result would be the 2×1 zero vector. Therefore, it is always the case that $A0=0$, for an appropriately sized zero matrix or vector.

2.1.4. The ij^{th} Entry of a Product

In previous sections, we used the entries of a matrix to describe the action of matrix addition and scalar multiplication. We can also study matrix multiplication using the entries of matrices.

What is the ij^{th} entry of AB? It is the entry in the i^{th} row and the j^{th} column of the product AB.

Now if A is $m \times n$ and B is $n \times p$, then we know that the product AB has the form

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nj} & \cdots & b_{np} \end{bmatrix}$$

The j^{th} column of AB is of the form

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

which is an $m \times 1$ column vector. It is calculated by

$$b_{1j} \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + b_{2j} \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + b_{nj} \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

Therefore, the ij^{th} entry is the entry in row i of this vector. This is computed by

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}$$

The following is the formal definition for the ij^{th} entry of a product of matrices.

Definition 2.21: The ij^{th} Entry of a Product

Let $A = [a_{ij}]$ be an $m \times n$ matrix and let $B = [b_{ij}]$ be an $n \times p$ matrix. Then AB is an $m \times p$ matrix and the (i, j)-entry of AB is defined as

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

Another way to write this is

$$(AB)_{ij} = \left[\begin{array}{cccc} a_{i1} & a_{i2} & \cdots & a_{in} \end{array} \right] \left[\begin{array}{c} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{array} \right] = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

In other words, to find the (i, j)-entry of the product AB, or $(AB)_{ij}$, you multiply the i^{th} row of A, on the left by the j^{th} column of B. To express AB in terms of its entries, we write $AB = [(AB)_{ij}]$.

Consider the following example.

Example 2.22: The Entries of a Product

Compute AB if possible. If it is, find the $(3, 2)$-entry of AB using Definition 2.21.

$$A = \left[\begin{array}{cc} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{array} \right], B = \left[\begin{array}{ccc} 2 & 3 & 1 \\ 7 & 6 & 2 \end{array} \right]$$

Solution. First check if the product is possible. It is of the form $(3 \times 2)(2 \times 3)$ and since the inside numbers match, it is possible to do the multiplication. The result should be a 3×3 matrix. We can first compute AB:

$$\left[\left[\begin{array}{cc} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{array} \right] \left[\begin{array}{c} 2 \\ 7 \end{array} \right], \left[\begin{array}{cc} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{array} \right] \left[\begin{array}{c} 3 \\ 6 \end{array} \right], \left[\begin{array}{cc} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{array} \right] \left[\begin{array}{c} 1 \\ 2 \end{array} \right] \right]$$

where the commas separate the columns in the resulting product. Thus the above product equals

$$\left[\begin{array}{rrr} 16 & 15 & 5 \\ 13 & 15 & 5 \\ 46 & 42 & 14 \end{array} \right]$$

which is a 3×3 matrix as desired. Thus, the $(3, 2)$-entry equals 42.

Now using Definition 2.21, we can find that the $(3,2)$-entry equals

$$\begin{aligned}\sum_{k=1}^{2} a_{3k}b_{k2} &= a_{31}b_{12} + a_{32}b_{22} \\ &= 2 \times 3 + 6 \times 6 = 42\end{aligned}$$

Consulting our result for AB above, this is correct!

You may wish to use this method to verify that the rest of the entries in AB are correct.

□

Here is another example.

Example 2.23: Finding the Entries of a Product

Determine if the product AB is defined. If it is, find the $(2,1)$-entry of the product.

$$A = \left[\begin{array}{ccc} 2 & 3 & 1 \\ 7 & 6 & 2 \\ 0 & 0 & 0 \end{array}\right], B = \left[\begin{array}{cc} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{array}\right]$$

Solution. This product is of the form $(3 \times 3)(3 \times 2)$. The middle numbers match so the matrices are conformable and it is possible to compute the product.

We want to find the $(2,1)$-entry of AB, that is, the entry in the second row and first column of the product. We will use Definition 2.21, which states

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

In this case, $n = 3$, $i = 2$ and $j = 1$. Hence the $(2,1)$-entry is found by computing

$$(AB)_{21} = \sum_{k=1}^{3} a_{2k}b_{k1} = \left[\begin{array}{ccc} a_{21} & a_{22} & a_{23} \end{array}\right] \left[\begin{array}{c} b_{11} \\ b_{21} \\ b_{31} \end{array}\right]$$

Substituting in the appropriate values, this product becomes

$$\left[\begin{array}{ccc} a_{21} & a_{22} & a_{23} \end{array}\right] \left[\begin{array}{c} b_{11} \\ b_{21} \\ b_{31} \end{array}\right] = \left[\begin{array}{ccc} 7 & 6 & 2 \end{array}\right] \left[\begin{array}{c} 1 \\ 3 \\ 2 \end{array}\right] = 1 \times 7 + 3 \times 6 + 2 \times 2 = 29$$

Hence, $(AB)_{21} = 29$.

You should take a moment to find a few other entries of AB. You can multiply the matrices to check that your answers are correct. The product AB is given by

$$AB = \left[\begin{array}{cc} 13 & 13 \\ 29 & 32 \\ 0 & 0 \end{array}\right]$$

□

2.1.5. Properties of Matrix Multiplication

As pointed out above, it is sometimes possible to multiply matrices in one order but not in the other order. However, even if both AB and BA are defined, they may not be equal.

Example 2.24: Matrix Multiplication is Not Commutative

Compare the products AB and BA, for matrices $A = \left[\begin{array}{cc} 1 & 2 \\ 3 & 4 \end{array}\right], B = \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right]$

Solution. First, notice that A and B are both of size 2×2. Therefore, both products AB and BA are defined. The first product, AB is

$$AB = \left[\begin{array}{cc} 1 & 2 \\ 3 & 4 \end{array}\right] \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right] = \left[\begin{array}{cc} 2 & 1 \\ 4 & 3 \end{array}\right]$$

The second product, BA is

$$\left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right] \left[\begin{array}{cc} 1 & 2 \\ 3 & 4 \end{array}\right] = \left[\begin{array}{cc} 3 & 4 \\ 1 & 2 \end{array}\right]$$

Therefore, $AB \neq BA$. ☐

This example illustrates that you cannot assume $AB = BA$ even when multiplication is defined in both orders. If for some matrices A and B it is true that $AB = BA$, then we say that A and B **commute**. This is one important property of matrix multiplication.

The following are other important properties of matrix multiplication. Notice that these properties hold only when the size of matrices are such that the products are defined.

Proposition 2.25: Properties of Matrix Multiplication

The following hold for matrices A, B, and C and for scalars r and s,

- $$A\left(rB + sC\right) = r\left(AB\right) + s\left(AC\right) \tag{2.6}$$
- $$\left(B + C\right)A = BA + CA \tag{2.7}$$
- $$A\left(BC\right) = \left(AB\right)C \tag{2.8}$$

Proof. First we will prove 2.6. We will use Definition 2.21 and prove this statement using the ij^{th} entries of a matrix. Therefore,

$$\left(A\left(rB + sC\right)\right)_{ij} = \sum_k a_{ik}\left(rB + sC\right)_{kj} = \sum_k a_{ik}\left(rb_{kj} + sc_{kj}\right)$$

$$= r\sum_k a_{ik}b_{kj} + s\sum_k a_{ik}c_{kj} = r\,(AB)_{ij} + s\,(AC)_{ij}$$
$$= (r\,(AB) + s\,(AC))_{ij}$$

Thus $A\,(rB + sC) = r(AB) + s(AC)$ as claimed.

The proof of 2.7 follows the same pattern and is left as an exercise.

Statement 2.8 is the associative law of multiplication. Using Definition 2.21,

$$(A\,(BC))_{ij} = \sum_k a_{ik}\,(BC)_{kj} = \sum_k a_{ik}\sum_l b_{kl}c_{lj}$$
$$= \sum_l (AB)_{il}\,c_{lj} = ((AB)\,C)_{ij}\,.$$

This proves 2.8. ☐

2.1.6. The Transpose

Another important operation on matrices is that of taking the **transpose**. For a matrix A, we denote the **transpose** of A by A^T. Before formally defining the transpose, we explore this operation on the following matrix.

$$\left[\begin{array}{cc} 1 & 4 \\ 3 & 1 \\ 2 & 6 \end{array}\right]^T = \left[\begin{array}{ccc} 1 & 3 & 2 \\ 4 & 1 & 6 \end{array}\right]$$

What happened? The first column became the first row and the second column became the second row. Thus the 3×2 matrix became a 2×3 matrix. The number 4 was in the first row and the second column and it ended up in the second row and first column.

The definition of the transpose is as follows.

Definition 2.26: The Transpose of a Matrix

Let A be an $m\times n$ matrix. Then A^T, the ***transpose*** *of A, denotes the $n\times m$ matrix given by*

$$A^T = [a_{ij}]^T = [a_{ji}]$$

The (i,j)-entry of A becomes the (j,i)-entry of A^T.

Consider the following example.

Example 2.27: The Transpose of a Matrix

Calculate A^T for the following matrix

$$A = \left[\begin{array}{rrr} 1 & 2 & -6 \\ 3 & 5 & 4 \end{array}\right]$$

Solution. By Definition 2.26, we know that for $A = [a_{ij}]$, $A^T = [a_{ji}]$. In other words, we switch the row and column location of each entry. The $(1,2)$-entry becomes the $(2,1)$-entry.

Thus,

$$A^T = \left[\begin{array}{rr} 1 & 3 \\ 2 & 5 \\ -6 & 4 \end{array}\right]$$

Notice that A is a 2×3 matrix, while A^T is a 3×2 matrix. ∎

The transpose of a matrix has the following important properties .

Lemma 2.28: Properties of the Transpose of a Matrix

Let A be an $m \times n$ matrix, B an $n \times p$ matrix, and r and s scalars. Then

1. $\left(A^T\right)^T = A$
2. $(AB)^T = B^T A^T$
3. $(rA + sB)^T = rA^T + sB^T$

Proof. First we prove 2. From Definition 2.26,

$$(AB)^T = [(AB)_{ij}]^T = [(AB)_{ji}] = \sum_k a_{jk} b_{ki} = \sum_k b_{ki} a_{jk}$$

$$= \sum_k [b_{ik}]^T [a_{kj}]^T = [b_{ij}]^T [a_{ij}]^T = B^T A^T$$

The proof of Formula 3 is left as an exercise. ∎

The transpose of a matrix is related to other important topics. Consider the following definition.

Definition 2.29: Symmetric and Skew Symmetric Matrices

An $n \times n$ matrix A is said to be ***symmetric*** *if $A = A^T$. It is said to be* ***skew symmetric*** *if $A = -A^T$.*

We will explore these definitions in the following examples.

Example 2.30: Symmetric Matrices

Let

$$A = \left[\begin{array}{rrr} 2 & 1 & 3 \\ 1 & 5 & -3 \\ 3 & -3 & 7 \end{array}\right]$$

Use Definition 2.29 to show that A is symmetric.

Solution. By Definition 2.29, we need to show that $A = A^T$. Now, using Definition 2.26,

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

Hence, $A = A^T$, so A is symmetric. ◻

Example 2.31: A Skew Symmetric Matrix

Let

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

Show that A is skew symmetric.

Solution. By Definition 2.29,

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

You can see that each entry of A^T is equal to -1 times the same entry of A. Hence, $A^T = -A$ and so by Definition 2.29, A is skew symmetric. ◻

2.1.7. The Identity and Inverses

There is a special matrix, denoted I, which is called to as the **identity matrix**. The identity matrix is always a square matrix, and it has the property that there are ones down the main diagonal and zeroes elsewhere. Here are some identity matrices of various sizes.

$$[1], \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The first is the 1×1 identity matrix, the second is the 2×2 identity matrix, and so on. By extension, you can likely see what the $n \times n$ identity matrix would be. When it is necessary to distinguish which size of identity matrix is being discussed, we will use the notation I_n for the $n \times n$ identity matrix.

The identity matrix is so important that there is a special symbol to denote the ij^{th} entry of the identity matrix. This symbol is given by $I_{ij} = \delta_{ij}$ where δ_{ij} is the **Kronecker symbol** defined by

$$\delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

I_n is called the **identity matrix** because it is a **multiplicative identity** in the following sense.

Lemma 2.32: Multiplication by the Identity Matrix

Suppose A is an $m \times n$ matrix and I_n is the $n \times n$ identity matrix. Then $AI_n = A$. If I_m is the $m \times m$ identity matrix, it also follows that $I_m A = A$.

Proof. The (i, j)-entry of AI_n is given by:

$$\sum_k a_{ik}\delta_{kj} = a_{ij}$$

and so $AI_n = A$. The other case is left as an exercise for you. $\square$

We now define the matrix operation which in some ways plays the role of division.

Definition 2.33: The Inverse of a Matrix

*A square $n \times n$ matrix A is said to have an **inverse** A^{-1} if and only if*

$$AA^{-1} = A^{-1}A = I_n$$

*In this case, the matrix A is called **invertible**.*

Such a matrix A^{-1} will have the same size as the matrix A. It is very important to observe that the inverse of a matrix, if it exists, is unique. Another way to think of this is that if it acts like the inverse, then it **is** the inverse.

Theorem 2.34: Uniqueness of Inverse

Suppose A is an $n \times n$ matrix such that an inverse A^{-1} exists. Then there is only one such inverse matrix. That is, given any matrix B such that $AB = BA = I$, $B = A^{-1}$.

Proof. In this proof, it is assumed that I is the $n \times n$ identity matrix. Let A, B be $n \times n$ matrices such that A^{-1} exists and $AB = BA = I$. We want to show that $A^{-1} = B$. Now using properties we have seen, we get:

$$A^{-1} = A^{-1}I = A^{-1}(AB) = \left(A^{-1}A\right)B = IB = B$$

Hence, $A^{-1} = B$ which tells us that the inverse is unique. $\square$

The next example demonstrates how to check the inverse of a matrix.

Example 2.35: Verifying the Inverse of a Matrix

Let $A = \left[\begin{array}{rr} 1 & 1 \\ 1 & 2 \end{array}\right]$. Show $\left[\begin{array}{rr} 2 & -1 \\ -1 & 1 \end{array}\right]$ is the inverse of A.

Solution. To check this, multiply

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

and

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

showing that this matrix is indeed the inverse of A. ☐

Unlike ordinary multiplication of numbers, it can happen that $A \neq 0$ but A may fail to have an inverse. This is illustrated in the following example.

Example 2.36: A Nonzero Matrix With No Inverse

Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. *Show that* A *does not have an inverse.*

Solution. One might think A would have an inverse because it does not equal zero. However, note that

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

If A^{-1} existed, we would have the following

$$\begin{aligned} \begin{bmatrix} 0 \\ 0 \end{bmatrix} &= A^{-1}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) \\ &= A^{-1}\left(A\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) \\ &= \left(A^{-1}A\right)\begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= I\begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ 1 \end{bmatrix} \end{aligned}$$

This says that

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

which is impossible! Therefore, A does not have an inverse. ☐

In the next section, we will explore how to find the inverse of a matrix, if it exists.

2.1.8. Finding the Inverse of a Matrix

In Example 2.35, we were given A^{-1} and asked to verify that this matrix was in fact the inverse of A. In this section, we explore how to find A^{-1}.

Let

$$A = \left[\begin{array}{cc} 1 & 1 \\ 1 & 2 \end{array} \right]$$

as in Example 2.35. In order to find A^{-1}, we need to find a matrix $\left[\begin{array}{cc} x & z \\ y & w \end{array} \right]$ such that

$$\left[\begin{array}{cc} 1 & 1 \\ 1 & 2 \end{array} \right] \left[\begin{array}{cc} x & z \\ y & w \end{array} \right] = \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right]$$

We can multiply these two matrices, and see that in order for this equation to be true, we must find the solution to the systems of equations,

$$\begin{array}{c} x + y = 1 \\ x + 2y = 0 \end{array}$$

and

$$\begin{array}{c} z + w = 0 \\ z + 2w = 1 \end{array}$$

Writing the augmented matrix for these two systems gives

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

for the first system and

$$\left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 2 & 1 \end{array} \right] \tag{2.9}$$

for the second.

Let's solve the first system. Take -1 times the first row and add to the second to get

$$\left[\begin{array}{cc|r} 1 & 1 & 1 \\ 0 & 1 & -1 \end{array} \right]$$

Now take -1 times the second row and add to the first to get

$$\left[\begin{array}{cc|r} 1 & 0 & 2 \\ 0 & 1 & -1 \end{array} \right]$$

Writing in terms of variables, this says $x = 2$ and $y = -1$.

Now solve the second system, 2.9 to find z and w. You will find that $z = -1$ and $w = 1$.

If we take the values found for x, y, z, and w and put them into our inverse matrix, we see that the inverse is

$$A^{-1} = \left[\begin{array}{cc} x & z \\ y & w \end{array} \right] = \left[\begin{array}{rr} 2 & -1 \\ -1 & 1 \end{array} \right]$$

After taking the time to solve the second system, you may have noticed that exactly the same row operations were used to solve both systems. In each case, the end result was something of the form $[I|X]$ where I is the identity and X gave a column of the inverse. In the above,

$$\left[\begin{array}{c} x \\ y \end{array}\right]$$

the first column of the inverse was obtained by solving the first system and then the second column

$$\left[\begin{array}{c} z \\ w \end{array}\right]$$

To simplify this procedure, we could have solved both systems at once! To do so, we could have written

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

and row reduced until we obtained

$$\left[\begin{array}{cc|rr} 1 & 0 & 2 & -1 \\ 0 & 1 & -1 & 1 \end{array}\right]$$

and read off the inverse as the 2×2 matrix on the right side.

This exploration motivates the following important algorithm.

Algorithm 2.37: Matrix Inverse Algorithm

Suppose A is an $n \times n$ matrix. To find A^{-1} if it exists, form the augmented $n \times 2n$ matrix

$$[A|I]$$

If possible do row operations until you obtain an $n \times 2n$ matrix of the form

$$[I|B]$$

When this has been done, $B = A^{-1}$. In this case, we say that A is ***invertible****. If it is impossible to row reduce to a matrix of the form $[I|B]$, then A has no inverse.*

This algorithm shows how to find the inverse if it exists. It will also tell you if A does not have an inverse.

Consider the following example.

Example 2.38: Finding the Inverse

Let $A = \left[\begin{array}{rrr} 1 & 2 & 2 \\ 1 & 0 & 2 \\ 3 & 1 & -1 \end{array}\right]$. *Find A^{-1} if it exists.*

Solution. Set up the augmented matrix

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

Now we row reduce, with the goal of obtaining the 3×3 identity matrix on the left hand side. First, take -1 times the first row and add to the second followed by -3 times the first row added to the third row. This yields

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & -2 & 0 & -1 & 1 & 0 \\ 0 & -5 & -7 & -3 & 0 & 1 \end{array}\right]$$

Then take 5 times the second row and add to -2 times the third row.

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & -10 & 0 & -5 & 5 & 0 \\ 0 & 0 & 14 & 1 & 5 & -2 \end{array}\right]$$

Next take the third row and add to -7 times the first row. This yields

$$\left[\begin{array}{ccc|ccc} -7 & -14 & 0 & -6 & 5 & -2 \\ 0 & -10 & 0 & -5 & 5 & 0 \\ 0 & 0 & 14 & 1 & 5 & -2 \end{array}\right]$$

Now take $-\frac{7}{5}$ times the second row and add to the first row.

$$\left[\begin{array}{ccc|ccc} -7 & 0 & 0 & 1 & -2 & -2 \\ 0 & -10 & 0 & -5 & 5 & 0 \\ 0 & 0 & 14 & 1 & 5 & -2 \end{array}\right]$$

Finally divide the first row by -7, the second row by -10 and the third row by 14 which yields

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{7} & \frac{2}{7} & \frac{2}{7} \\ 0 & 1 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 & \frac{1}{14} & \frac{5}{14} & -\frac{1}{7} \end{array}\right]$$

Notice that the left hand side of this matrix is now the 3×3 identity matrix I_3. Therefore, the inverse is the 3×3 matrix on the right hand side, given by

$$\left[\begin{array}{ccc} -\frac{1}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{14} & \frac{5}{14} & -\frac{1}{7} \end{array}\right]$$

$\square$

It may happen that through this algorithm, you discover that the left hand side cannot be row reduced to the identity matrix. Consider the following example of this situation.

Example 2.39: A Matrix Which Has No Inverse

Let $A = \left[\begin{array}{rrr} 1 & 2 & 2 \\ 1 & 0 & 2 \\ 2 & 2 & 4 \end{array} \right]$. *Find* A^{-1} *if it exists.*

Solution. Write the augmented matrix $\left[A|I \right]$

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

and proceed to do row operations attempting to obtain $\left[I|A^{-1}\right]$. Take -1 times the first row and add to the second. Then take -2 times the first row and add to the third row.

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

Next add -1 times the second row to the third row.

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

At this point, you can see there will be no way to obtain I on the left side of this augmented matrix. Hence, there is no way to complete this algorithm, and therefore the inverse of A does not exist. In this case, we say that A is not invertible. $\square$

If the algorithm provides an inverse for the original matrix, it is always possible to check your answer. To do so, use the method demonstrated in Example 2.35. Check that the products AA^{-1} and $A^{-1}A$ both equal the identity matrix. Through this method, you can always be sure that you have calculated A^{-1} properly!

One way in which the inverse of a matrix is useful is to find the solution of a system of linear equations. Recall from Definition 2.15 that we can write a system of equations in matrix form, which is of the form $AX = B$. Suppose you find the inverse of the matrix A^{-1}. Then you could multiply both sides of this equation on the left by A^{-1} and simplify to obtain

$$\begin{aligned} \left(A^{-1}\right) AX &= A^{-1}B \\ \left(A^{-1}A\right) X &= A^{-1}B \\ IX &= A^{-1}B \\ X &= A^{-1}B \end{aligned}$$

Therefore we can find X, the solution to the system, by computing $X = A^{-1}B$. Note that once you have found A^{-1}, you can easily get the solution for different right hand sides (different B). It is always just $A^{-1}B$.

We will explore this method of finding the solution to a system in the following example.

Example 2.40: Using the Inverse to Solve a System of Equations

Consider the following system of equations. Use the inverse of a suitable matrix to give the solutions to this system.

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

Solution. First, we can write the system of equations in matrix form

$$AX = \left[\begin{array}{rrr} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{array}\right] \left[\begin{array}{r} x \\ y \\ z \end{array}\right] = \left[\begin{array}{r} 1 \\ 3 \\ 2 \end{array}\right] = B \tag{2.10}$$

The inverse of the matrix

$$A = \left[\begin{array}{rrr} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{array}\right]$$

is

$$A^{-1} = \left[\begin{array}{rrr} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{array}\right]$$

Verifying this inverse is left as an exercise.

From here, the solution to the given system 2.10 is found by

$$\left[\begin{array}{r} x \\ y \\ z \end{array}\right] = A^{-1}B = \left[\begin{array}{rrr} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{array}\right] \left[\begin{array}{r} 1 \\ 3 \\ 2 \end{array}\right] = \left[\begin{array}{r} \frac{5}{2} \\ -2 \\ -\frac{3}{2} \end{array}\right]$$

♠

What if the right side, B, of 2.10 had been $\left[\begin{array}{r} 0 \\ 1 \\ 3 \end{array}\right]$? In other words, what would be the solution to

$$\left[\begin{array}{rrr} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{array}\right] \left[\begin{array}{r} x \\ y \\ z \end{array}\right] = \left[\begin{array}{r} 0 \\ 1 \\ 3 \end{array}\right]?$$

By the above discussion, the solution is given by

$$\left[\begin{array}{c} x \\ y \\ z \end{array}\right] = A^{-1}B = \left[\begin{array}{rrr} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{array}\right] \left[\begin{array}{c} 0 \\ 1 \\ 3 \end{array}\right] = \left[\begin{array}{r} 2 \\ -1 \\ -2 \end{array}\right]$$

This illustrates that for a system $AX = B$ where A^{-1} exists, it is easy to find the solution when the vector B is changed.

We conclude this section with some important properties of the inverse.

Theorem 2.41: Inverses of Transposes and Products

Let A, B, and A_i for $i = 1, ..., k$ be $n \times n$ matrices.

1. *If A is an invertible matrix, then $(A^T)^{-1} = (A^{-1})^T$*
2. *If A and B are invertible matrices, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$*
3. *If $A_1, A_2, ..., A_k$ are invertible, then the product $A_1A_2 \cdots A_k$ is invertible, and $(A_1A_2 \cdots A_k)^{-1} = A_k^{-1}A_{k-1}^{-1} \cdots A_2^{-1}A_1^{-1}$*

Consider the following theorem.

Theorem 2.42: Properties of the Inverse

Let A be an $n \times n$ matrix and I the usual identity matrix.

1. *I is invertible and $I^{-1} = I$*
2. *If A is invertible then so is A^{-1}, and $(A^{-1})^{-1} = A$*
3. *If A is invertible then so is A^k, and $(A^k)^{-1} = (A^{-1})^k$*
4. *If A is invertible and p is a nonzero real number, then pA is invertible and $(pA)^{-1} = \frac{1}{p}A^{-1}$*

2.1.9. Elementary Matrices

We now turn our attention to a special type of matrix called an **elementary matrix**. An elementary matrix is always a square matrix. Recall the row operations given in Definition 1.11. Any elementary matrix, which we often denote by E, is obtained from applying *one* row operation to the identity matrix of the same size.

For example, the matrix

$$E = \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right]$$

is the elementary matrix obtained from switching the two rows. The matrix

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

is the elementary matrix obtained from multiplying the second row of the 3×3 identity matrix by 3. The matrix

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

is the elementary matrix obtained from adding -3 times the first row to the third row.

You may construct an elementary matrix from any row operation, but remember that you can only apply one operation.

Consider the following definition.

Definition 2.43: Elementary Matrices and Row Operations

Let E be an $n \times n$ matrix. Then E is an ***elementary matrix*** *if it is the result of applying one row operation to the $n \times n$ identity matrix I_n.*
Those which involve switching rows of the identity matrix are called permutation matrices.

Therefore, E constructed above by switching the two rows of I_2 is called a permutation matrix.

Elementary matrices can be used in place of row operations and therefore are very useful. It turns out that multiplying (on the left hand side) by an elementary matrix E will have the same effect as doing the row operation used to obtain E.

The following theorem is an important result which we will use throughout this text.

Theorem 2.44: Multiplication by an Elementary Matrix and Row Operations

To perform any of the three row operations on a matrix A it suffices to take the product EA, where E is the elementary matrix obtained by using the desired row operation on the identity matrix.

Therefore, instead of performing row operations on a matrix A, we can row reduce through matrix multiplication with the appropriate elementary matrix. We will examine this theorem in detail for each of the three row operations given in Definition 1.11.

First, consider the following lemma.

Lemma 2.45: Action of Permutation Matrix

Let P^{ij} denote the elementary matrix which involves switching the i^{th} and the j^{th} rows. Then P^{ij} is a permutation matrix and

$$P^{ij}A = B$$

where B is obtained from A by switching the i^{th} and the j^{th} rows.

We will explore this idea more in the following example.

Example 2.46: Switching Rows with an Elementary Matrix

Let

$$P^{12} = \left[\begin{array}{ccc} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{array} \right], A = \left[\begin{array}{cc} a & b \\ g & d \\ e & f \end{array} \right]$$

Find B where $B = P^{12}A$.

Solution. You can see that the matrix P^{12} is obtained by switching the first and second rows of the 3×3 identity matrix I.

Using our usual procedure, compute the product $P^{12}A = B$. The result is given by

$$B = \left[\begin{array}{cc} g & d \\ a & b \\ e & f \end{array} \right]$$

Notice that B is the matrix obtained by switching rows 1 and 2 of A. Therefore by multiplying A by P^{12}, the row operation which was applied to I to obtain P^{12} is applied to A to obtain B. ∎

Theorem 2.44 applies to all three row operations, and we now look at the row operation of multiplying a row by a scalar. Consider the following lemma.

Lemma 2.47: Multiplication by a Scalar and Elementary Matrices

Let $E(k,i)$ denote the elementary matrix corresponding to the row operation in which the i^{th} row is multiplied by the nonzero scalar, k. Then

$$E(k,i)A = B$$

where B is obtained from A by multiplying the i^{th} row of A by k.

We will explore this lemma further in the following example.

Example 2.48: Multiplication of a Row by 5 Using Elementary Matrix

Let

$$E(5,2) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$$

Find the matrix B where $B = E(5,2)A$

Solution. You can see that $E(5,2)$ is obtained by multiplying the second row of the identity matrix by 5.

Using our usual procedure for multiplication of matrices, we can compute the product $E(5,2)A$. The resulting matrix is given by

$$B = \begin{bmatrix} a & b \\ 5c & 5d \\ e & f \end{bmatrix}$$

Notice that B is obtained by multiplying the second row of A by the scalar 5. ☐

There is one last row operation to consider. The following lemma discusses the final operation of adding a multiple of a row to another row.

Lemma 2.49: Adding Multiples of Rows and Elementary Matrices

Let $E(k \times i + j)$ denote the elementary matrix obtained from I by adding k times the i^{th} row to the j^{th}. Then

$$E(k \times i + j)A = B$$

where B is obtained from A by adding k times the i^{th} row to the j^{th} row of A.

Consider the following example.

Example 2.50: Adding Two Times the First Row to the Last

Let

$$E(2 \times 1 + 3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}, A = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$$

Find B where $B = E(2 \times 1 + 3)A$.

Solution. You can see that the matrix $E(2 \times 1 + 3)$ was obtained by adding 2 times the first row of I to the third row of I.

Using our usual procedure, we can compute the product $E(2 \times 1 + 3)A$. The resulting matrix B is given by

$$B = \begin{bmatrix} a & b \\ c & d \\ 2a + e & 2b + f \end{bmatrix}$$

You can see that B is the matrix obtained by adding 2 times the first row of A to the third row. ◻

Suppose we have applied a row operation to a matrix A. Consider the row operation required to return A to its original form, to undo the row operation. It turns out that this action is how we find the inverse of an elementary matrix E.

Consider the following theorem.

Theorem 2.51: Elementary Matrices and Inverses

Every elementary matrix is invertible and its inverse is also an elementary matrix.

In fact, the inverse of an elementary matrix is constructed by doing the *reverse* row operation on I. E^{-1} will be obtained by performing the row operation which would carry E back to I.

- If E is obtained by switching rows i and j, then E^{-1} is also obtained by switching rows i and j.
- If E is obtained by multiplying row i by the scalar k, then E^{-1} is obtained by multiplying row i by the scalar $\frac{1}{k}$.
- If E is obtained by adding k times row i to row j, then E^{-1} is obtained by subtracting k times row i from row j.

Consider the following example.

Example 2.52: Inverse of an Elementary Matrix

Let

$$E = \left[\begin{array}{cc} 1 & 0 \\ 0 & 2 \end{array} \right]$$

Find E^{-1}.

Solution. Consider the elementary matrix E given by

$$E = \left[\begin{array}{cc} 1 & 0 \\ 0 & 2 \end{array} \right]$$

Here, E is obtained from the 2×2 identity matrix by multiplying the second row by 2. In order to carry E back to the identity, we need to multiply the second row of E by $\frac{1}{2}$. Hence, E^{-1} is given by

$$E^{-1} = \left[\begin{array}{cc} 1 & 0 \\ 0 & \frac{1}{2} \end{array} \right]$$

We can verify that $EE^{-1} = I$. Take the product EE^{-1}, given by

$$EE^{-1} = \left[\begin{array}{cc} 1 & 0 \\ 0 & 2 \end{array} \right] \left[\begin{array}{cc} 1 & 0 \\ 0 & \frac{1}{2} \end{array} \right] = \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right]$$

This equals I so we know that we have compute E^{-1} properly. ◻

Suppose an $m \times n$ matrix A is row reduced to its reduced row-echelon form. By tracking each row operation completed, this row reduction can be completed through multiplication by elementary matrices. Consider the following definition.

Definition 2.53: The Form $B = UA$

Let A be an $m \times n$ matrix and let B be the reduced row-echelon form of A. Then we can write $B = UA$ where U is the product of all elementary matrices representing the row operations done to A to obtain B.

Consider the following example.

Example 2.54: The Form $B = UA$

Let $A = \left[\begin{array}{rr} 0 & 1 \\ 1 & 0 \\ 2 & 0 \end{array} \right]$. Find B, the reduced row-echelon form of A and write it in the form $B = UA$.

Solution. To find B, row reduce A. For each step, we will record the appropriate elementary matrix. First, switch rows 1 and 2.

$$\left[\begin{array}{rr} 0 & 1 \\ 1 & 0 \\ 2 & 0 \end{array} \right] \rightarrow \left[\begin{array}{rr} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{array} \right]$$

The resulting matrix is equivalent to finding the product of $P^{12} = \left[\begin{array}{rrr} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{array} \right]$ and A.

Next, add (-2) times row 1 to row 3.

$$\left[\begin{array}{rr} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{array} \right] \rightarrow \left[\begin{array}{rr} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{array} \right]$$

This is equivalent to multiplying by the matrix $E(-2 \times 1 + 3) = \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{array} \right]$. Notice that the resulting matrix is B, the required reduced row-echelon form of A.

We can then write

$$\begin{array}{rcl} B & = & E(-2 \times 1 + 2) \left(P^{12} A \right) \\ & = & \left(E(-2 \times 1 + 2) P^{12} \right) A \\ & = & UA \end{array}$$

It remains to find the matrix U.

$$\begin{aligned} U &= E(-2\times 1+2)P^{12} \\ &= \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{array}\right]\left[\begin{array}{rrr} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{array}\right] \\ &= \left[\begin{array}{rrr} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & -2 & 1 \end{array}\right] \end{aligned}$$

We can verify that $B = UA$ holds for this matrix U:

$$\begin{aligned} UA &= \left[\begin{array}{rrr} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & -2 & 1 \end{array}\right]\left[\begin{array}{rr} 0 & 1 \\ 1 & 0 \\ 2 & 0 \end{array}\right] \\ &= \left[\begin{array}{rr} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{array}\right] \\ &= B \end{aligned}$$

$\square$

While the process used in the above example is reliable and simple when only a few row operations are used, it becomes cumbersome in a case where many row operations are needed to carry A to B. The following theorem provides an alternate way to find the matrix U.

Theorem 2.55: Finding the Matrix U

Let A be an $m \times n$ matrix and let B be its reduced row-echelon form. Then $B = UA$ where U is an invertible $m \times m$ matrix found by forming the matrix $[A|I_m]$ and row reducing to $[B|U]$.

Let's revisit the above example using the process outlined in Theorem 2.55.

Example 2.56: The Form $B = UA$, Revisited

Let $A = \left[\begin{array}{rr} 0 & 1 \\ 1 & 0 \\ 2 & 0 \end{array}\right]$. Using the process outlined in Theorem 2.55, find U such that $B = UA$.

Solution. First, set up the matrix $[A|I_m]$.

$$\left[\begin{array}{rr|rrr} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & 0 & 1 \end{array}\right]$$

Now, row reduce this matrix until the left side equals the reduced row-echelon form of A.

$$\left[\begin{array}{cc|ccc} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & 0 & 1 \end{array}\right] \rightarrow \left[\begin{array}{cc|ccc} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 \end{array}\right]$$
$$\rightarrow \left[\begin{array}{cc|ccc} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -2 & 1 \end{array}\right]$$

The left side of this matrix is B, and the right side is U. Comparing this to the matrix U found above in Example 2.54, you can see that the same matrix is obtained regardless of which process is used. ▯

Recall from Algorithm 2.37 that an $n \times n$ matrix A is invertible if and only if A can be carried to the $n \times n$ identity matrix using the usual row operations. This leads to an important consequence related to the above discussion.

Suppose A is an $n \times n$ invertible matrix. Then, set up the matrix $[A|I_n]$ as done above, and row reduce until it is of the form $[B|U]$. In this case, $B = I_n$ because A is invertible.

$$\begin{aligned} B &= UA \\ I_n &= UA \\ U^{-1} &= A \end{aligned}$$

Now suppose that $U = E_1 E_2 \cdots E_k$ where each E_i is an elementary matrix representing a row operation used to carry A to I. Then,

$$U^{-1} = (E_1 E_2 \cdots E_k)^{-1} = E_k^{-1} \cdots E_2^{-1} E_1 {-1}$$

Remember that if E_i is an elementary matrix, so too is E_i^{-1}. It follows that

$$\begin{aligned} A &= U^{-1} \\ &= E_k^{-1} \cdots E_2^{-1} E_1 {-1} \end{aligned}$$

and A can be written as a product of elementary matrices.

Theorem 2.57: Product of Elementary Matrices

Let A be an $n \times n$ matrix. Then A is invertible if and only if it can be written as a product of elementary matrices.

Consider the following example.

Example 2.58: Product of Elementary Matrices

Let $A = \left[\begin{array}{ccc} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & -2 & 1 \end{array}\right]$. *Write A as a product of elementary matrices.*

Solution. We will use the process outlined in Theorem 2.55 to write A as a product of elementary matrices. We will set up the matrix $[A|I]$ and row reduce, recording each row operation as an elementary matrix.

First:

$$\left[\begin{array}{rrr|rrr} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & -2 & 1 & 0 & 0 & 1 \end{array}\right] \rightarrow \left[\begin{array}{rrr|rrr} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 & 0 & 1 \end{array}\right]$$

represented by the elementary matrix $E_1 = \left[\begin{array}{rrr} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{array}\right]$.

Secondly:

$$\left[\begin{array}{rrr|rrr} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 & 0 & 1 \end{array}\right] \rightarrow \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 & 0 & 1 \end{array}\right]$$

represented by the elementary matrix $E_2 = \left[\begin{array}{rrr} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$.

Finally:

$$\left[\begin{array}{rrr|rrr} 1 & 0 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 & 0 & 1 \end{array}\right] \rightarrow \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 1 \end{array}\right]$$

represented by the elementary matrix $E_3 = \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{array}\right]$.

Notice that the reduced row-echelon form of A is I. Hence $I = UA$ where U is the product of the above elementary matrices. It follows that $A = U^{-1}$. Since we want to write A as a product of elementary matrices, we wish to express U^{-1} as a product of elementary matrices.

$$\begin{aligned} U^{-1} &= (E_3E_2E_1)^{-1} \\ &= E_1^{-1}E_2^{-1}E_3^{-1} \\ &= \left[\begin{array}{rrr} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{array}\right]\left[\begin{array}{rrr} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]\left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{array}\right] \\ &= A \end{aligned}$$

This gives A written as a product of elementary matrices. By Theorem 2.57 it follows that A is invertible. ∎

2.1.10. More on Matrix Inverses

In this section, we will prove three theorems which will clarify the concept of matrix inverses. In order to do this, first recall some important properties of elementary matrices.

Recall that an elementary matrix is a square matrix obtained by performing an elementary operation on an identity matrix. Each elementary matrix is invertible, and its inverse is also an elementary matrix. If E is an $m \times m$ elementary matrix and A is an $m \times n$ matrix, then the product EA is the result of applying to A the same elementary row operation that was applied to the $m \times m$ identity matrix in order to obtain E.

Let R be the reduced row-echelon form of an $m \times n$ matrix A. R is obtained by iteratively applying a sequence of elementary row operations to A. Denote by $E_1, E_2, \cdots, E_k$ the elementary matrices associated with the elementary row operations which were applied, in order, to the matrix A to obtain the resulting R. We then have that $R = (E_k \cdots (E_2 (E_1 A))) = E_k \cdots E_2 E_1 A$. Let E denote the product matrix $E_k \cdots E_2 E_1$ so that we can write $R = EA$ where E is an invertible matrix whose inverse is the product $(E_1)^{-1}(E_2)^{-1} \cdots (E_k)^{-1}$.

Now, we will consider some preliminary lemmas.

Lemma 2.59: Invertible Matrix and Zeros

Suppose that A and B are matrices such that the product AB is an identity matrix. Then the reduced row-echelon form of A does not have a row of zeros.

Proof: Let R be the reduced row-echelon form of A. Then $R = EA$ for some invertible square matrix E as described above. By hypothesis $AB = I$ where I is an identity matrix, so we have a chain of equalities

$$R(BE^{-1}) = (EA)(BE^{-1}) = E(AB)E^{-1} = EIE^{-1} = EE^{-1} = I$$

If R would have a row of zeros, then so would the product $R(BE^{-1})$. But since the identity matrix I does not have a row of zeros, neither can R have one. ■

We now consider a second important lemma.

Lemma 2.60: Size of Invertible Matrix

Suppose that A and B are matrices such that the product AB is an identity matrix. Then A has at least as many columns as it has rows.

Proof: Let R be the reduced row-echelon form of A. By Lemma 2.59, we know that R does not have a row of zeros, and therefore each row of R has a leading 1. Since each column of R contains at most one of these leading 1s, R must have at least as many columns as it has rows. ■

An important theorem follows from this lemma.

Theorem 2.61: Invertible Matrices are Square

Only square matrices can be invertible.

Proof: Suppose that A and B are matrices such that both products AB and BA are identity matrices. We will show that A and B must be square matrices of the same size. Let the matrix A have m rows and n columns, so that A is an $m \times n$ matrix. Since the product

AB exists, B must have n rows, and since the product BA exists, B must have m columns so that B is an $n \times m$ matrix. To finish the proof, we need only verify that $m = n$.

We first apply Lemma 2.60 with A and B, to obtain the inequality $m \leq n$. We then apply Lemma 2.60 again (switching the order of the matrices), to obtain the inequality $n \leq m$. It follows that $m = n$, as we wanted. ■

Of course, not all square matrices are invertible. In particular, zero matrices are not invertible, along with many other square matrices.

The following proposition will be useful in proving the next theorem.

Proposition 2.62: Reduced Row-Echelon Form of a Square Matrix

If R is the reduced row-echelon form of a square matrix, then either R has a row of zeros or R is an identity matrix.

The proof of this proposition is left as an exercise to the reader. We now consider the second important theorem of this section.

Theorem 2.63: Unique Inverse of a Matrix

Suppose A and B are square matrices such that $AB = I$ where I is an identity matrix. Then it follows that $BA = I$. Further, both A and B are invertible and $B = A^{-1}$ and $A = B^{-1}$.

Proof: Let R be the reduced row-echelon form of a square matrix A. Then, $R = EA$ where E is an invertible matrix. Since $AB = I$, Lemma 2.59 gives us that R does not have a row of zeros. By noting that R is a square matrix and applying Proposition 2.62, we see that $R = I$. Hence, $EA = I$.

Using both that $EA = I$ and $AB = I$, we can finish the proof with a chain of equalities as given by

$$\begin{aligned} BA = IBIA &= (EA)B(E^{-1}E)A \\ &= E(AB)E^{-1}(EA) \\ &= EIE^{-1}I \\ &= EE^{-1} = I \end{aligned}$$

It follows from the definition of the inverse of a matrix that $B = A^{-1}$ and $A = B^{-1}$. ■

This theorem is very useful, since with it we need only test one of the products AB or BA in order to check that B is the inverse of A. The hypothesis that A and B are square matrices is very important, and without this the theorem does not hold.

We will now consider an example.

Example 2.64: Non Square Matrices

Let

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

Show that $A^T A = I$ *but* $AA^T \neq 0$.

Solution. Consider the product $A^T A$ given by

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Therefore, $A^T A = I_2$, where I_2 is the 2×2 identity matrix. However, the product AA^T is

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence AA^T is not the 3×3 identity matrix. This shows that for Theorem 2.63, it is essential that both matrices be square and of the same size. ☐

Is it possible to have matrices A and B such that $AB = I$, while $BA = 0$? This question is left to the reader to answer, and you should take a moment to consider the answer.

We conclude this section with an important theorem.

Theorem 2.65: The Reduced Row-Echelon Form of an Invertible Matrix

For any matrix A the following conditions are equivalent:

- *A is invertible*
- *The reduced row-echelon form of A is an identity matrix*

Proof. In order to prove this, we show that for any given matrix A, each condition implies the other. We first show that if A is invertible, then its reduced row-echelon form is an identity matrix, then we show that if the reduced row-echelon form of A is an identity matrix, then A is invertible.

If A is invertible, there is some matrix B such that $AB = I$. By Lemma 2.59, we get that the reduced row-echelon form of A does not have a row of zeros. Then by Theorem 2.61, it follows that A and the reduced row-echelon form of A are square matrices. Finally, by Proposition 2.62, this reduced row-echelon form of A must be an identity matrix. This proves the first implication.

Now suppose the reduced row-echelon form of A is an identity matrix I. Then $I = EA$ for some product E of elementary matrices. By Theorem 2.63, we can conclude that A is invertible. ☐

Theorem 2.65 corresponds to Algorithm 2.37, which claims that A^{-1} is found by row reducing the augmented matrix $[A|I]$ to the form $[I|A^{-1}]$. This will be a matrix product $E\,[A|I]$ where E is a product of elementary matrices. By the rules of matrix multiplication, we have that $E\,[A|I] = [EA|EI] = [EA|E]$.

It follows that the reduced row-echelon form of $[A|I]$ is $[EA|E]$, where EA gives the reduced row-echelon form of A. By Theorem 2.65, if $EA \neq I$, then A is not invertible, and if $EA = I$, A is invertible. If $EA = I$, then by Theorem 2.63, $E = A^{-1}$. This proves that Algorithm 2.37 does in fact find A^{-1}.

Exercises

Exercise 2.1.1 *For the following pairs of matrices, determine if the sum $A+B$ is defined. If so, find the sum.*

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

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

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

Exercise 2.1.2 *For each matrix A, find the matrix $-A$ such that $A+(-A)=0$.*

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

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

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

Exercise 2.1.3 *In the context of Proposition 2.7, describe $-A$ and 0.*

Exercise 2.1.4 *For each matrix A, find the product $(-2)A, 0A,$ and $3A$.*

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

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

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

Exercise 2.1.5 *Using only the properties given in Proposition 2.7 and Proposition 2.10, show* $-A$ *is unique.*

Exercise 2.1.6 *Using only the properties given in Proposition 2.7 and Proposition 2.10 show* 0 *is unique.*

Exercise 2.1.7 *Using only the properties given in Proposition 2.7 and Proposition 2.10 show* $0A = 0$*. Here the* 0 *on the left is the scalar* 0 *and the* 0 *on the right is the zero matrix of appropriate size.*

Exercise 2.1.8 *Using only the properties given in Proposition 2.7 and Proposition 2.10, as well as previous problems show* $(-1)\,A = -A$*.*

Exercise 2.1.9 *Consider the matrices* $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 7 \end{bmatrix}, B = \begin{bmatrix} 3 & -1 & 2 \\ -3 & 2 & 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix},$
$D = \begin{bmatrix} -1 & 2 \\ 2 & -3 \end{bmatrix}, E = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$

Find the following if possible. If it is not possible explain why.

1. $-3A$
2. $3B - A$
3. AC
4. CB
5. AE
6. EA

Exercise 2.1.10 *Consider the matrices* $A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \\ 1 & -1 \end{bmatrix}, B = \begin{bmatrix} 2 & -5 & 2 \\ -3 & 2 & 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 2 \\ 5 & 0 \end{bmatrix},$
$D = \begin{bmatrix} -1 & 1 \\ 4 & -3 \end{bmatrix}, E = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

Find the following if possible. If it is not possible explain why.

1. $-3A$
2. $3B - A$
3. AC
4. CA

5. AE

6. EA

7. BE

8. DE

Exercise 2.1.11 *Let* $A = \begin{bmatrix} 1 & 1 \\ -2 & -1 \\ 1 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 1 & -1 & -2 \\ 2 & 1 & -2 \end{bmatrix}$, *and* $C = \begin{bmatrix} 1 & 1 & -3 \\ -1 & 2 & 0 \\ -3 & -1 & 0 \end{bmatrix}$. *Find the following if possible.*

1. AB

2. BA

3. AC

4. CA

5. CB

6. BC

Exercise 2.1.12 *Let* $A = \begin{bmatrix} -1 & -1 \\ 3 & 3 \end{bmatrix}$. *Find all* 2×2 *matrices,* B *such that* $AB = 0$.

Exercise 2.1.13 *Let* $X = \begin{bmatrix} -1 & -1 & 1 \end{bmatrix}$ *and* $Y = \begin{bmatrix} 0 & 1 & 2 \end{bmatrix}$. *Find* $X^T Y$ *and* XY^T *if possible.*

Exercise 2.1.14 *Let* $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 & 2 \\ 3 & k \end{bmatrix}$. *Is it possible to choose* k *such that* $AB = BA$? *If so, what should* k *equal?*

Exercise 2.1.15 *Let* $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 & 2 \\ 1 & k \end{bmatrix}$. *Is it possible to choose* k *such that* $AB = BA$? *If so, what should* k *equal?*

Exercise 2.1.16 *Find* 2×2 *matrices,* A, B, *and* C *such that* $A \neq 0, C \neq B$, *but* $AC = AB$.

Exercise 2.1.17 *Give an example of matrices (of any size),* A, B, C *such that* $B \neq C$, $A \neq 0$, *and yet* $AB = AC$.

Exercise 2.1.18 *Find* 2×2 *matrices* A *and* B *such that* $A \neq 0$ *and* $B \neq 0$ *but* $AB = 0$.

Exercise 2.1.19 *Give an example of matrices (of any size),* A, B *such that* $A \neq 0$ *and* $B \neq 0$ *but* $AB = 0$.

Exercise 2.1.20 *Find* 2×2 *matrices* A *and* B *such that* $A \neq 0$ *and* $B \neq 0$ *with* $AB \neq BA$.

Exercise 2.1.21 *Write the system*

$$\begin{gathered} x_1 - x_2 + 2x_3 \\ 2x_3 + x_1 \\ 3x_3 \\ 3x_4 + 3x_2 + x_1 \end{gathered}$$

in the form $A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ *where* A *is an appropriate matrix.*

Exercise 2.1.22 *Write the system*

$$\begin{gathered} x_1 + 3x_2 + 2x_3 \\ 2x_3 + x_1 \\ 6x_3 \\ x_4 + 3x_2 + x_1 \end{gathered}$$

in the form $A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ *where* A *is an appropriate matrix.*

Exercise 2.1.23 *Write the system*

$$\begin{gathered} x_1 + x_2 + x_3 \\ 2x_3 + x_1 + x_2 \\ x_3 - x_1 \\ 3x_4 + x_1 \end{gathered}$$

in the form $A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ *where* A *is an appropriate matrix.*

Exercise 2.1.24 *A matrix* A *is called* idempotent *if* $A^2 = A$*. Let*

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

and show that A *is idempotent .*

Exercise 2.1.25 *For each pair of matrices, find the* $(1,2)$*-entry and* $(2,3)$*-entry of the product* AB*.*

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

2. $A = \begin{bmatrix} 1 & 3 & 1 \\ 0 & 2 & 4 \\ 1 & 0 & 5 \end{bmatrix}, B = \begin{bmatrix} 2 & 3 & 0 \\ -4 & 16 & 1 \\ 0 & 2 & 2 \end{bmatrix}$

Exercise 2.1.26 *Suppose A and B are square matrices of the same size. Which of the following are necessarily true?*

1. $(A-B)^2 = A^2 - 2AB + B^2$
2. $(AB)^2 = A^2B^2$
3. $(A+B)^2 = A^2 + 2AB + B^2$
4. $(A+B)^2 = A^2 + AB + BA + B^2$
5. $A^2B^2 = A(AB)B$
6. $(A+B)^3 = A^3 + 3A^2B + 3AB^2 + B^3$
7. $(A+B)(A-B) = A^2 - B^2$

Exercise 2.1.27 *Consider the matrices* $A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \\ 1 & -1 \end{bmatrix}, B = \begin{bmatrix} 2 & -5 & 2 \\ -3 & 2 & 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 2 \\ 5 & 0 \end{bmatrix}$,
$D = \begin{bmatrix} -1 & 1 \\ 4 & -3 \end{bmatrix}, E = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

Find the following if possible. If it is not possible explain why.

1. $-3A^T$
2. $3B - A^T$
3. E^TB
4. EE^T
5. B^TB
6. CA^T
7. D^TBE

Exercise 2.1.28 *Let A be an $n \times n$ matrix. Show A equals the sum of a symmetric and a skew symmetric matrix.* ***Hint:*** *Show that $\frac{1}{2}\left(A^T + A\right)$ is symmetric and then consider using this as one of the matrices.*

Exercise 2.1.29 *Show that the main diagonal of every skew symmetric matrix consists of only zeros. Recall that the main diagonal consists of every entry of the matrix which is of the form* a_{ii}.

Exercise 2.1.30 *Prove 3. That is, show that for an* $m \times n$ *matrix* A, *an* $n \times p$ *matrix* B, *and scalars* r, s, *the following holds:*

$$(rA + sB)^T = rA^T + sB^T$$

Exercise 2.1.31 *Prove that* $I_m A = A$ *where* A *is an* $m \times n$ *matrix.*

Exercise 2.1.32 *Suppose* $AB = AC$ *and* A *is an invertible* $n \times n$ *matrix. Does it follow that* $B = C$? *Explain why or why not.*

Exercise 2.1.33 *Suppose* $AB = AC$ *and* A *is a non invertible* $n \times n$ *matrix. Does it follow that* $B = C$? *Explain why or why not.*

Exercise 2.1.34 *Give an example of a matrix* A *such that* $A^2 = I$ *and yet* $A \neq I$ *and* $A \neq -I$.

Exercise 2.1.35 *Let*

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

Find A^{-1} *if possible. If* A^{-1} *does not exist, explain why.*

Exercise 2.1.36 *Let*

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

Find A^{-1} *if possible. If* A^{-1} *does not exist, explain why.*

Exercise 2.1.37 *Let*

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

Find A^{-1} *if possible. If* A^{-1} *does not exist, explain why.*

Exercise 2.1.38 *Let*

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

Find A^{-1} *if possible. If* A^{-1} *does not exist, explain why.*

Exercise 2.1.39 *Let* A *be a* 2×2 *invertible matrix, with* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. *Find a formula for* A^{-1} *in terms of* a, b, c, d.

Exercise 2.1.40 *Let*

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

Find A^{-1} if possible. If A^{-1} does not exist, explain why.

Exercise 2.1.41 *Let*

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

Find A^{-1} if possible. If A^{-1} does not exist, explain why.

Exercise 2.1.42 *Let*

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 4 & 5 & 10 \end{bmatrix}$$

Find A^{-1} if possible. If A^{-1} does not exist, explain why.

Exercise 2.1.43 *Let*

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

Find A^{-1} if possible. If A^{-1} does not exist, explain why.

Exercise 2.1.44 *Using the inverse of the matrix, find the solution to the systems:*

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

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

Now give the solution in terms of a and b to

$$\begin{bmatrix} 2 & 4 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

Exercise 2.1.45 *Using the inverse of the matrix, find the solution to the systems:*

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

2.

$$\begin{bmatrix} 1 & 0 & 3 \\ 2 & 3 & 4 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix}$$

Now give the solution in terms of a, b, and c to the following:

$$\begin{bmatrix} 1 & 0 & 3 \\ 2 & 3 & 4 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Exercise 2.1.46 *Show that if A is an $n \times n$ invertible matrix and X is a $n \times 1$ matrix such that $AX = B$ for B an $n \times 1$ matrix, then $X = A^{-1}B$.*

Exercise 2.1.47 *Prove that if A^{-1} exists and $AX = 0$ then $X = 0$.*

Exercise 2.1.48 *Show that if A^{-1} exists for an $n \times n$ matrix, then it is unique. That is, if $BA = I$ and $AB = I$, then $B = A^{-1}$.*

Exercise 2.1.49 *Show that if A is an invertible $n \times n$ matrix, then so is A^T and $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.*

Exercise 2.1.50 *Show $(AB)^{-1} = B^{-1}A^{-1}$ by verifying that*

$$AB\left(B^{-1}A^{-1}\right) = I$$

and

$$B^{-1}A^{-1}(AB) = I$$

Hint: *Use Problem 2.1.48.*

Exercise 2.1.51 *Show that $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$ by verifying that $(ABC)\left(C^{-1}B^{-1}A^{-1}\right) = I$ and $\left(C^{-1}B^{-1}A^{-1}\right)(ABC) = I$.* ***Hint:*** *Use Problem 2.1.48.*

Exercise 2.1.52 *If A is invertible, show $\left(A^2\right)^{-1} = \left(A^{-1}\right)^2$.* ***Hint:*** *Use Problem 2.1.48.*

Exercise 2.1.53 *If A is invertible, show $\left(A^{-1}\right)^{-1} = A$.* ***Hint:*** *Use Problem 2.1.48.*

Exercise 2.1.54 *Let $A = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$. Suppose a row operation is applied to A and the result is $B = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$. Find the elementary matrix E that represents this row operation.*

Exercise 2.1.55 *Let $A = \begin{bmatrix} 4 & 0 \\ 2 & 1 \end{bmatrix}$. Suppose a row operation is applied to A and the result is $B = \begin{bmatrix} 8 & 0 \\ 2 & 1 \end{bmatrix}$. Find the elementary matrix E that represents this row operation.*

Exercise 2.1.56 *Let* $A = \begin{bmatrix} 1 & -3 \\ 0 & 5 \end{bmatrix}$. *Suppose a row operation is applied to* A *and the result is* $B = \begin{bmatrix} 1 & -3 \\ 2 & -1 \end{bmatrix}$. *Find the elementary matrix* E *that represents this row operation.*

Exercise 2.1.57 *Let* $A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 5 & 1 \\ 2 & -1 & 4 \end{bmatrix}$. *Suppose a row operation is applied to* A *and the result is* $B = \begin{bmatrix} 1 & 2 & 1 \\ 2 & -1 & 4 \\ 0 & 5 & 1 \end{bmatrix}$.

1. *Find the elementary matrix* E *such that* $EA = B$.
2. *Find the inverse of* E, E^{-1}, *such that* $E^{-1}B = A$.

Exercise 2.1.58 *Let* $A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 5 & 1 \\ 2 & -1 & 4 \end{bmatrix}$. *Suppose a row operation is applied to* A *and the result is* $B = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 10 & 2 \\ 2 & -1 & 4 \end{bmatrix}$.

1. *Find the elementary matrix* E *such that* $EA = B$.
2. *Find the inverse of* E, E^{-1}, *such that* $E^{-1}B = A$.

Exercise 2.1.59 *Let* $A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 5 & 1 \\ 2 & -1 & 4 \end{bmatrix}$. *Suppose a row operation is applied to* A *and the result is* $B = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 5 & 1 \\ 1 & -\frac{1}{2} & 2 \end{bmatrix}$.

1. *Find the elementary matrix* E *such that* $EA = B$.
2. *Find the inverse of* E, E^{-1}, *such that* $E^{-1}B = A$.

Exercise 2.1.60 *Let* $A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 5 & 1 \\ 2 & -1 & 4 \end{bmatrix}$. *Suppose a row operation is applied to* A *and the result is* $B = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 5 \\ 2 & -1 & 4 \end{bmatrix}$.

1. *Find the elementary matrix* E *such that* $EA = B$.
2. *Find the inverse of* E, E^{-1}, *such that* $E^{-1}B = A$.

3. Determinants

3.1 Basic Techniques and Properties

Outcomes

A. *Evaluate the determinant of a square matrix using either Laplace Expansion or row operations.*

B. *Demonstrate the effects that row operations have on determinants.*

C. *Verify the following:*

  (a) *The determinant of a product of matrices is the product of the determinants.*

  (b) *The determinant of a matrix is equal to the determinant of its transpose.*

3.1.1. Cofactors and 2×2 Determinants

Let A be an $n \times n$ matrix. That is, let A be a square matrix. The **determinant** of A, denoted by $\det(A)$ is a very important number which we will explore throughout this section.

If A is a 2×2 matrix, the determinant is given by the following formula.

Definition 3.1: Determinant of a Two By Two Matrix

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. *Then*

$$\det(A) = ad - cb$$

The determinant is also often denoted by enclosing the matrix with two vertical lines. Thus

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

The following is an example of finding the determinant of a 2×2 matrix.

Example 3.2: A Two by Two Determinant

Find $\det(A)$ *for the matrix* $A = \left[\begin{array}{rr} 2 & 4 \\ -1 & 6 \end{array}\right]$.

Solution. From Definition 3.1,

$$\det(A) = (2)(6) - (-1)(4) = 12 + 4 = 16$$

□

The 2×2 determinant can be used to find the determinant of larger matrices. We will now explore how to find the determinant of a 3×3 matrix, using several tools including the 2×2 determinant.

We begin with the following definition.

Definition 3.3: The ij^{th} Minor of a Matrix

Let A be a 3×3 matrix. The ij^{th} ***minor*** *of A, denoted as $minor(A)_{ij}$, is the determinant of the 2×2 matrix which results from deleting the i^{th} row and the j^{th} column of A.*

In general, if A is an $n \times n$ matrix, then the ij^{th} minor of A is the determinant of the $n-1 \times n-1$ matrix which results from deleting the i^{th} row and the j^{th} column of A.

Hence, there is a minor associated with each entry of A. Consider the following example which demonstrates this definition.

Example 3.4: Finding Minors of a Matrix

Let

$$A = \left[\begin{array}{rrr} 1 & 2 & 3 \\ 4 & 3 & 2 \\ 3 & 2 & 1 \end{array}\right]$$

Find $minor(A)_{12}$ and $minor(A)_{23}$.

Solution. First we will find $minor(A)_{12}$. By Definition 3.3, this is the determinant of the 2×2 matrix which results when you delete the first row and the second column. This minor is given by

$$minor(A)_{12} = \det \left[\begin{array}{rr} 4 & 2 \\ 3 & 1 \end{array}\right]$$

Using Definition 3.1, we see that

$$\det \left[\begin{array}{rr} 4 & 2 \\ 3 & 1 \end{array}\right] = (4)(1) - (3)(2) = 4 - 6 = -2$$

Therefore $minor(A)_{12} = -2$.

Similarly, $minor(A)_{23}$ is the determinant of the 2×2 matrix which results when you delete the second row and the third column. This minor is therefore

$$minor(A)_{23} = \det \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} = -4$$

Finding the other minors of A is left as an exercise. ☐

The ij^{th} minor of a matrix A is used in another important definition, given next.

Definition 3.5: The ij^{th} Cofactor of a Matrix

*Suppose A is an $n \times n$ matrix. The ij^{th} **cofactor**, denoted by $cof(A)_{ij}$ is defined to be*

$$cof(A)_{ij} = (-1)^{i+j}\, minor(A)_{ij}$$

It is also convenient to refer to the cofactor of an entry of a matrix as follows. If a_{ij} is the ij^{th} entry of the matrix, then its cofactor is just $\operatorname{cof}(A)_{ij}$.

Example 3.6: Finding Cofactors of a Matrix

Consider the matrix

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

Find $cof(A)_{12}$ and $cof(A)_{23}$.

Solution. We will use Definition 3.5 to compute these cofactors.

First, we will compute $\operatorname{cof}(A)_{12}$. Therefore, we need to find $minor(A)_{12}$. This is the determinant of the 2×2 matrix which results when you delete the first row and the second column. Thus $minor(A)_{12}$ is given by

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

Then,

$$\operatorname{cof}(A)_{12} = (-1)^{1+2}\, minor(A)_{12} = (-1)^{1+2}(-2) = 2$$

Hence, $\operatorname{cof}(A)_{12} = 2$.

Similarly, we can find $\operatorname{cof}(A)_{23}$. First, find $minor(A)_{23}$, which is the determinant of the 2×2 matrix which results when you delete the second row and the third column. This minor is therefore

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

Hence,

$$\operatorname{cof}(A)_{23} = (-1)^{2+3}\, minor(A)_{23} = (-1)^{2+3}(-4) = 4$$

$\square$

You may wish to find the remaining cofactors for the above matrix. Remember that there is a cofactor for every entry in the matrix.

We have now established the tools we need to find the determinant of a 3×3 matrix.

Definition 3.7: The Determinant of a Three By Three Matrix

Let A be a 3×3 matrix. Then, $\det(A)$ is calculated by picking a row (or column) and taking the product of each entry in that row (column) with its cofactor and adding these products together.
*This process when applied to the i^{th} row (column) is known as **expanding along the i^{th} row (column)** as is given by*

$$\det(A) = a_{i1}\mathit{cof}(A)_{i1} + a_{i2}\mathit{cof}(A)_{i2} + a_{i3}\mathit{cof}(A)_{i3}$$

When calculating the determinant, you can choose to expand any row or any column. Regardless of your choice, you will always get the same number which is the determinant of the matrix A. This method of evaluating a determinant by expanding along a row or a column is called **Laplace Expansion** or **Cofactor Expansion**.

Consider the following example.

Example 3.8: Finding the Determinant of a Three by Three Matrix

Let

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

Find $\det(A)$ using the method of Laplace Expansion.

Solution. First, we will calculate $\det(A)$ by expanding along the first column. Using Definition 3.7, we take the 1 in the first column and multiply it by its cofactor,

$$1\left(-1\right)^{1+1}\begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix} = (1)(1)(-1) = -1$$

Similarly, we take the 4 in the first column and multiply it by its cofactor, as well as with the 3 in the first column. Finally, we add these numbers together, as given in the following equation.

$$\det(A) = 1\overbrace{(-1)^{1+1}\begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{11}} + 4\overbrace{(-1)^{2+1}\begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{21}} + 3\overbrace{(-1)^{3+1}\begin{vmatrix} 2 & 3 \\ 3 & 2 \end{vmatrix}}^{\mathrm{cof}(A)_{31}}$$

Calculating each of these, we obtain

$$\det(A) = 1\left(1\right)\left(-1\right) + 4\left(-1\right)\left(-4\right) + 3\left(1\right)\left(-5\right) = -1 + 16 + -15 = 0$$

Hence, $\det(A) = 0$.

As mentioned in Definition 3.7, we can choose to expand along any row or column. Let's try now by expanding along the second row. Here, we take the 4 in the second row and multiply it to its cofactor, then add this to the 3 in the second row multiplied by its cofactor, and the 2 in the second row multiplied by its cofactor. The calculation is as follows.

$$\det(A) = 4\overbrace{(-1)^{2+1}\left|\begin{array}{cc} 2 & 3 \\ 2 & 1 \end{array}\right|}^{\operatorname{cof}(A)_{21}} + 3\overbrace{(-1)^{2+2}\left|\begin{array}{cc} 1 & 3 \\ 3 & 1 \end{array}\right|}^{\operatorname{cof}(A)_{22}} + 2\overbrace{(-1)^{2+3}\left|\begin{array}{cc} 1 & 2 \\ 3 & 2 \end{array}\right|}^{\operatorname{cof}(A)_{23}}$$

Calculating each of these products, we obtain

$$\det(A) = 4(-1)(-2) + 3(1)(-8) + 2(-1)(-4) = 0$$

You can see that for both methods, we obtained $\det(A) = 0$. $\square$

As mentioned above, we will always come up with the same value for $\det(A)$ regardless of the row or column we choose to expand along. You should try to compute the above determinant by expanding along other rows and columns. This is a good way to check your work, because you should come up with the same number each time!

We present this idea formally in the following theorem.

Theorem 3.9: The Determinant is Well Defined

Expanding the $n \times n$ matrix along any row or column always gives the same answer, which is the determinant.

We have now looked at the determinant of 2×2 and 3×3 matrices. It turns out that the method used to calculate the determinant of a 3×3 matrix can be used to calculate the determinant of any sized matrix. Notice that Definition 3.3, Definition 3.5 and Definition 3.7 can all be applied to a matrix of any size.

For example, the ij^{th} minor of a 4×4 matrix is the determinant of the 3×3 matrix you obtain when you delete the i^{th} row and the j^{th} column. Just as with the 3×3 determinant, we can compute the determinant of a 4×4 matrix by Laplace Expansion, along any row or column

Consider the following example.

Example 3.10: Determinant of a Four by Four Matrix

Find $\det(A)$ *where*

$$A = \left[\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 5 & 4 & 2 & 3 \\ 1 & 3 & 4 & 5 \\ 3 & 4 & 3 & 2 \end{array}\right]$$

Solution. As in the case of a 3×3 matrix, you can expand this along any row or column. Lets pick the third column. Then, using Laplace Expansion,

$$\det(A) = 3(-1)^{1+3} \begin{vmatrix} 5 & 4 & 3 \\ 1 & 3 & 5 \\ 3 & 4 & 2 \end{vmatrix} + 2(-1)^{2+3} \begin{vmatrix} 1 & 2 & 4 \\ 1 & 3 & 5 \\ 3 & 4 & 2 \end{vmatrix} +$$

$$4(-1)^{3+3} \begin{vmatrix} 1 & 2 & 4 \\ 5 & 4 & 3 \\ 3 & 4 & 2 \end{vmatrix} + 3(-1)^{4+3} \begin{vmatrix} 1 & 2 & 4 \\ 5 & 4 & 3 \\ 1 & 3 & 5 \end{vmatrix}$$

Now, you can calculate each 3×3 determinant using Laplace Expansion, as we did above. You should complete these as an exercise and verify that $\det(A) = -12$. ☐

The following provides a formal definition for the determinant of an $n \times n$ matrix. You may wish to take a moment and consider the above definitions for 2×2 and 3×3 determinants in context of this definition.

Definition 3.11: The Determinant of an $n \times n$ Matrix

Let A be an $n \times n$ matrix where $n \geq 2$ and suppose the determinant of an $(n-1) \times (n-1)$ has been defined. Then

$$\det(A) = \sum_{j=1}^{n} a_{ij} \, cof(A)_{ij} = \sum_{i=1}^{n} a_{ij} \, cof(A)_{ij}$$

The first formula consists of expanding the determinant along the i^{th} row and the second expands the determinant along the j^{th} column.

In the following sections, we will explore some important properties and characteristics of the determinant.

3.1.2. The Determinant of a Triangular Matrix

There is a certain type of matrix for which finding the determinant is a very simple procedure. Consider the following definition.

Definition 3.12: Triangular Matrices

A matrix A is upper triangular if $a_{ij} = 0$ whenever $i > j$. Thus the entries of such a matrix below the main diagonal equal 0, as shown. Here, $*$ refers to any nonzero number.

$$\left[\begin{array}{cccc} * & * & \cdots & * \\ 0 & * & \cdots & \vdots \\ \vdots & \vdots & \ddots & * \\ 0 & \cdots & 0 & * \end{array} \right]$$

A lower triangular matrix is defined similarly as a matrix for which all entries above the main diagonal are equal to zero.

The following theorem provides a useful way to calculate the determinant of a triangular matrix.

Theorem 3.13: Determinant of a Triangular Matrix

Let A be an upper or lower triangular matrix. Then $\det(A)$ is obtained by taking the product of the entries on the main diagonal.

The verification of this Theorem can be done by computing the determinant using Laplace Expansion along the first row or column.

Consider the following example.

Example 3.14: Determinant of a Triangular Matrix

Let

$$A = \left[\begin{array}{rrrr} 1 & 2 & 3 & 77 \\ 0 & 2 & 6 & 7 \\ 0 & 0 & 3 & 33.7 \\ 0 & 0 & 0 & -1 \end{array} \right]$$

Find $\det(A)$.

Solution. From Theorem 3.13, it suffices to take the product of the elements on the main diagonal. Thus $\det(A) = 1 \times 2 \times 3 \times (-1) = -6$.

Without using Theorem 3.13, you could use Laplace Expansion. We will expand along the first column. This gives

$$\begin{aligned} \det(A) = \quad & 1 \left| \begin{array}{rrr} 2 & 6 & 7 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{array} \right| + 0(-1)^{2+1} \left| \begin{array}{rrr} 2 & 3 & 77 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{array} \right| + \\ & 0(-1)^{3+1} \left| \begin{array}{rrr} 2 & 3 & 77 \\ 2 & 6 & 7 \\ 0 & 0 & -1 \end{array} \right| + 0(-1)^{4+1} \left| \begin{array}{rrr} 2 & 3 & 77 \\ 2 & 6 & 7 \\ 0 & 3 & 33.7 \end{array} \right| \end{aligned}$$

and the only nonzero term in the expansion is

$$1\begin{vmatrix} 2 & 6 & 7 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix}$$

Now find the determinant of this 3×3 matrix, by expanding along the first column to obtain

$$\begin{aligned} \det(A) &= 1 \times \left(2 \times \begin{vmatrix} 3 & 33.7 \\ 0 & -1 \end{vmatrix} + 0\,(-1)^{2+1} \begin{vmatrix} 6 & 7 \\ 0 & -1 \end{vmatrix} + 0\,(-1)^{3+1} \begin{vmatrix} 6 & 7 \\ 3 & 33.7 \end{vmatrix} \right) \\ &= 1 \times 2 \times \begin{vmatrix} 3 & 33.7 \\ 0 & -1 \end{vmatrix} \end{aligned}$$

Next use Definition 3.1 to find the determinant of this 2×2 matrix, which is just $3 \times -1 - 0 \times 33.7 = -3$. Putting all these steps together, we have

$$\det(A) = 1 \times 2 \times 3 \times (-1) = -6$$

which is just the product of the entries down the main diagonal of the original matrix! ☐

You can see that while both methods result in the same answer, Theorem 3.13 provides a much quicker method.

In the next section, we explore some important properties of determinants.

3.1.3. Properties of Determinants I: Examples

There are many important properties of determinants. Since many of these properties involve the row operations discussed in Chapter 1, we recall that definition now.

Definition 3.15: Row Operations

The row operations consist of the following

1. *Switch two rows.*
2. *Multiply a row by a nonzero number.*
3. *Replace a row by a multiple of another row added to itself.*

We will now consider the effect of row operations on the determinant of a matrix. In future sections, we will see that using the following properties can greatly assist in finding determinants. This section will use the theorems as motivation to provide various examples of the usefulness of the properties.

The first theorem explains the affect on the determinant of a matrix when two rows are switched.

Theorem 3.16: Switching Rows

Let A be an $n \times n$ matrix and let B be a matrix which results from switching two rows of A. Then $\det(B) = -\det(A)$.

When we switch two rows of a matrix, the determinant is multiplied by -1. Consider the following example.

Example 3.17: Switching Two Rows

Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ *and let* $B = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}$. *Knowing that* $\det(A) = -2$, *find* $\det(B)$.

Solution. By Definition 3.1, $\det(A) = 1 \times 4 - 3 \times 2 = -2$. Notice that the rows of B are the rows of A but switched. By Theorem 3.16 since two rows of A have been switched, $\det(B) = -\det(A) = -(-2) = 2$. You can verify this using Definition 3.1. ☐

The next theorem demonstrates the effect on the determinant of a matrix when we multiply a row by a scalar.

Theorem 3.18: Multiplying a Row by a Scalar

Let A be an $n \times n$ matrix and let B be a matrix which results from multiplying some row of A by a scalar k. Then $\det(B) = k \det(A)$.

Notice that this theorem is true when we multiply *one* row of the matrix by k. If we were to multiply *two* rows of A by k to obtain B, we would have $\det(B) = k^2 \det(A)$. Suppose we were to multiply all n rows of A by k to obtain the matrix B, so that $B = kA$. Then, $\det(B) = k^n \det(A)$. This gives the next theorem.

Theorem 3.19: Scalar Multiplication

Let A and B be $n \times n$ matrices and k a scalar, such that $B = kA$. Then $\det(B) = k^n \det(A)$.

Consider the following example.

Example 3.20: Multiplying a Row by 5

Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 5 & 10 \\ 3 & 4 \end{bmatrix}$. *Knowing that* $\det(A) = -2$, *find* $\det(B)$.

Solution. By Definition 3.1, $\det(A) = -2$. We can also compute $\det(B)$ using Definition 3.1, and we see that $\det(B) = -10$.

Now, let's compute $\det(B)$ using Theorem 3.18 and see if we obtain the same answer. Notice that the first row of B is 5 times the first row of A, while the second row of B is equal to the second row of A. By Theorem 3.18, $\det(B) = 5 \times \det(A) = 5 \times -2 = -10$.

You can see that this matches our answer above. ☐

Finally, consider the next theorem for the last row operation, that of adding a multiple of a row to another row.

Theorem 3.21: Adding a Multiple of a Row to Another Row

Let A be an $n \times n$ matrix and let B be a matrix which results from adding a multiple of a row to another row. Then $\det(A) = \det(B)$.

Therefore, when we add a multiple of a row to another row, the determinant of the matrix is unchanged. Note that if a matrix A contains a row which is a multiple of another row, $\det(A)$ will equal 0. To see this, suppose the first row of A is equal to -1 times the second row. By Theorem 3.21, we can add the first row to the second row, and the determinant will be unchanged. However, this row operation will result in a row of zeros. Using Laplace Expansion along the row of zeros, we find that the determinant is 0.

Consider the following example.

Example 3.22: Adding a Row to Another Row

Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ *and let* $B = \begin{bmatrix} 1 & 2 \\ 5 & 8 \end{bmatrix}$. *Find* $\det(B)$.

Solution. By Definition 3.1, $\det(A) = -2$. Notice that the second row of B is two times the first row of A added to the second row. By Theorem 3.16, $\det(B) = \det(A) = -2$. As usual, you can verify this answer using Definition 3.1. ☐

Example 3.23: Multiple of a Row

Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. *Show that* $\det(A) = 0$.

Solution. Using Definition 3.1, the determinant is given by

$$\det(A) = 1 \times 4 - 2 \times 2 = 0$$

However notice that the second row is equal to 2 times the first row. Then by the discussion above following Theorem 3.21 the determinant will equal 0. ☐

Until now, our focus has primarily been on row operations. However, we can carry out the same operations with columns, rather than rows. The three operations outlined in Definition 3.15 can be done with columns instead of rows. In this case, in Theorems 3.16, 3.18, and 3.21 you can replace the word, "row" with the word "column".

There are several other major properties of determinants which do not involve row (or column) operations. The first is the determinant of a product of matrices.

Theorem 3.24: Determinant of a Product

Let A and B be two $n \times n$ matrices. Then

$$\det(AB) = \det(A)\det(B)$$

In order to find the determinant of a product of matrices, we can simply take the product of the determinants.

Consider the following example.

Example 3.25: The Determinant of a Product

Compare $\det(AB)$ *and* $\det(A)\det(B)$ *for*

$$A = \left[\begin{array}{rr} 1 & 2 \\ -3 & 2 \end{array}\right], B = \left[\begin{array}{rr} 3 & 2 \\ 4 & 1 \end{array}\right]$$

Solution. First compute AB, which is given by

$$AB = \left[\begin{array}{rr} 1 & 2 \\ -3 & 2 \end{array}\right]\left[\begin{array}{rr} 3 & 2 \\ 4 & 1 \end{array}\right] = \left[\begin{array}{rr} 11 & 4 \\ -1 & -4 \end{array}\right]$$

and so by Definition 3.1

$$\det(AB) = \det\left[\begin{array}{rr} 11 & 4 \\ -1 & -4 \end{array}\right] = -40$$

Now

$$\det(A) = \det\left[\begin{array}{rr} 1 & 2 \\ -3 & 2 \end{array}\right] = 8$$

and

$$\det(B) = \det\left[\begin{array}{rr} 3 & 2 \\ 4 & 1 \end{array}\right] = -5$$

Computing $\det(A) \times \det(B)$ we have $8 \times -5 = -40$. This is the same answer as above and you can see that $\det(A)\det(B) = 8 \times (-5) = -40 = \det(AB)$. ♠

Consider the next important property.

Theorem 3.26: Determinant of the Transpose

Let A be a matrix where A^T is the transpose of A. Then,

$$\det\left(A^T\right) = \det(A)$$

This theorem is illustrated in the following example.

Example 3.27: Determinant of the Transpose

Let

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

Find $\det\left(A^T\right)$.

Solution. First, note that

$$A^T = \begin{bmatrix} 2 & 4 \\ 5 & 3 \end{bmatrix}$$

Using Definition 3.1, we can compute $\det(A)$ and $\det\left(A^T\right)$. It follows that $\det(A) = 2 \times 3 - 4 \times 5 = -14$ and $\det\left(A^T\right) = 2 \times 3 - 5 \times 4 = -14$. Hence, $\det(A) = \det\left(A^T\right)$. ◻

The following provides an essential property of the determinant, as well as a useful way to determine if a matrix is invertible.

Theorem 3.28: Determinant of the Inverse

Let A be an $n \times n$ matrix. Then A is invertible if and only if $\det(A) \neq 0$. If this is true, it follows that

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

Consider the following example.

Example 3.29: Determinant of an Invertible Matrix

Let $A = \begin{bmatrix} 3 & 6 \\ 2 & 4 \end{bmatrix}, B = \begin{bmatrix} 2 & 3 \\ 5 & 1 \end{bmatrix}$. *For each matrix, determine if it is invertible. If so, find the determinant of the inverse.*

Solution. Consider the matrix A first. Using Definition 3.1 we can find the determinant as follows:

$$\det(A) = 3 \times 4 - 2 \times 6 = 12 - 12 = 0$$

By Theorem 3.28 A is not invertible.

Now consider the matrix B. Again by Definition 3.1 we have

$$\det(B) = 2 \times 1 - 5 \times 3 = 2 - 15 = -13$$

By Theorem 3.28 B is invertible and the determinant of the inverse is given by

$$\begin{aligned} \det\left(A^{-1}\right) &= \frac{1}{\det(A)} \\ &= \frac{1}{-13} \\ &= -\frac{1}{13} \end{aligned}$$

$\square$

3.1.4. Properties of Determinants II: Some Important Proofs

This section includes some important proofs on determinants and cofactors.

First we recall the definition of a determinant. If $A = [a_{ij}]$ is an $n \times n$ matrix, then $\det A$ is defined by computing the expansion along the first row:

$$\det A = \sum_{i=1}^{n} a_{1,i} \operatorname{cof}(A)_{1,i}. \tag{3.1}$$

If $n = 1$ then $\det A = a_{1,1}$.

The following example is straightforward and strongly recommended as a means for getting used to definitions.

Example 3.30

(1) Let E_{ij} be the elementary matrix obtained by interchanging ith and jth rows of I. Then $\det E_{ij} = -1$.
(2) Let E_{ik} be the elementary matrix obtained by multiplying the ith row of I by k. Then $\det E_{ik} = k$.
(3) Let E_{ijk} be the elementary matrix obtained by multiplying ith row of I by k and adding it to its jth row. Then $\det E_{ijk} = 1$.
(4) If C and B are such that CB is defined and the ith row of C consists of zeros, then the ith row of CB consists of zeros.
(5) If E is an elementary matrix, then $\det E = \det E^T$.

Many of the proofs in section use the Principle of Mathematical Induction. This concept is discussed in Appendix A.2 and is reviewed here for convenience. First we check that the assertion is true for $n = 2$ (the case $n = 1$ is either completely trivial or meaningless).

Next, we assume that the assertion is true for $n - 1$ (where $n \geq 3$) and prove it for n. Once this is accomplished, by the Principle of Mathematical Induction we can conclude that the statement is true for all $n \times n$ matrices for every $n \geq 2$.

If A is an $n \times n$ matrix and $1 \leq j \leq n$, then the matrix obtained by removing 1st column and jth row from A is an $n - 1 \times n - 1$ matrix (we shall denote this matrix by $A(j)$ below). Since these matrices are used in computation of cofactors $\operatorname{cof}(A)_{1,i}$, for $1 \leq i \neq n$, the inductive assumption applies to these matrices.

Consider the following lemma.

Lemma 3.31

If A is an $n \times n$ matrix such that one of its rows consists of zeros, then $\det A = 0$.

Proof. We will prove this lemma using Mathematical Induction.

If $n = 2$ this is easy (check!).

Let $n \geq 3$ be such that every matrix of size $n-1 \times n-1$ with a row consisting of zeros has determinant equal to zero. Let i be such that the ith row of A consists of zeros. Then we have $a_{ij} = 0$ for $1 \leq j \leq n$.

Fix $j \in \{1, 2, \ldots, n\}$ such that $j \neq i$. Then matrix $A(j)$ used in computation of $\text{cof}(A)_{1,j}$ has a row consisting of zeros, and by our inductive assumption $\text{cof}(A)_{1,j} = 0$.

On the other hand, if $j = i$ then $a_{1,j} = 0$. Therefore $a_{1,j}\text{cof}(A)_{1,j} = 0$ for all j and by (3.1) we have

$$\det A = \sum_{j=1}^{n} a_{1,j}\text{cof}(A)_{1,j} = 0$$

as each of the summands is equal to 0. □

Lemma 3.32

Assume A, B and C are $n \times n$ matrices that for some $1 \leq i \leq n$ satisfy the following.

1. *jth rows of all three matrices are identical, for $j \neq i$.*
2. *Each entry in the jth row of A is the sum of the corresponding entries in jth rows of B and C.*

Then $\det A = \det B + \det C$.

Proof. This is not difficult to check for $n = 2$ (do check it!).

Now assume that the statement of Lemma is true for $n-1 \times n-1$ matrices and fix A, B and C as in the statement. The assumptions state that we have $a_{l,j} = b_{l,j} = c_{l,j}$ for $j \neq i$ and for $1 \leq l \leq n$ and $a_{l,i} = b_{l,i} + c_{l,i}$ for all $1 \leq l \leq n$. Therefore $A(i) = B(i) = C(i)$, and $A(j)$ has the property that its ith row is the sum of ith rows of $B(j)$ and $C(j)$ for $j \neq i$ while the other rows of all three matrices are identical. Therefore by our inductive assumption we have $\text{cof}(A)_{1j} = \text{cof}(B)_{1j} + \text{cof}(C)_{1j}$ for $j \neq i$.

By (3.1) we have (using all equalities established above)

$$\begin{aligned}
\det A &= \sum_{l=1}^{n} a_{1,l}\text{cof}(A)_{1,l} \\
&= \sum_{l \neq i} a_{1,l}(\text{cof}(B)_{1,l} + \text{cof}(C)_{1,l}) + (b_{1,i} + c_{1,i})\text{cof}(A)_{1,i} \\
&= \det B + \det C
\end{aligned}$$

This proves that the assertion is true for all n and completes the proof. □

Theorem 3.33

Let A and B be $n \times n$ matrices.

1. *If A is obtained by interchanging ith and jth rows of B (with $i \neq j$), then* $\det A = -\det B$.

2. *If A is obtained by multiplying ith row of B by k then* $\det A = k \det B$.

3. *If two rows of A are identical then* $\det A = 0$.

4. *If A is obtained by multiplying ith row of B by k and adding it to jth row of B ($i \neq j$) then* $\det A = \det B$.

Proof. We prove all statements by induction. The case $n = 2$ is easily checked directly (and it is strongly suggested that you do check it).

We assume $n \geq 3$ and (1)–(4) are true for all matrices of size $n - 1 \times n - 1$.

(1) We prove the case when $j = i + 1$, i.e., we are interchanging two consecutive rows.

Let $l \in \{1, \dots, n\} \setminus \{i, j\}$. Then $A(l)$ is obtained from $B(l)$ by interchanging two of its rows (draw a picture) and by our assumption

$$\mathrm{cof}(A)_{1,l} = -\mathrm{cof}(B)_{1,l}. \tag{3.2}$$

Now consider $a_{1,i}\mathrm{cof}(A)_{1,l}$. We have that $a_{1,i} = b_{1,j}$ and also that $A(i) = B(j)$. Since $j = i + 1$, we have

$$(-1)^{1+j} = (-1)^{1+i+1} = -(-1)^{1+i}$$

and therefore $a_{1i}\mathrm{cof}(A)_{1i} = -b_{1j}\mathrm{cof}(B)_{1j}$ and $a_{1j}\mathrm{cof}(A)_{1j} = -b_{1i}\mathrm{cof}(B)_{1i}$. Putting this together with (3.2) into (3.1) we see that if in the formula for $\det A$ we change the sign of each of the summands we obtain the formula for $\det B$.

$$\det A = \sum_{l=1}^{n} a_{1l}\mathrm{cof}(A)_{1l} = -\sum_{l=1}^{n} b_{1l}B_{1l} = \det B.$$

We have therefore proved the case of (1) when $j = i + 1$. In order to prove the general case, one needs the following fact. If $i < j$, then in order to interchange ith and jth row one can proceed by interchanging two adjacent rows $2(j - i) + 1$ times: First swap ith and $i + 1$st, then $i + 1$st and $i + 2$nd, and so on. After one interchanges $j - 1$st and jth row, we have ith row in position of jth and lth row in position of $l - 1$st for $i + 1 \leq l \leq j$. Then proceed backwards swapping adjacent rows until everything is in place.

Since $2(j - i) + 1$ is an odd number $(-1)^{2(j-i)+1} = -1$ and we have that $\det A = -\det B$.

(2) This is like (1)… but much easier. Assume that (2) is true for all $n-1 \times n-1$ matrices. We have that $a_{ji} = kb_{ji}$ for $1 \leq j \leq n$. In particular $a_{1i} = kb_{1i}$, and for $l \neq i$ matrix $A(l)$ is obtained from $B(l)$ by multiplying one of its rows by k. Therefore $\mathrm{cof}(A)_{1l} = k\mathrm{cof}(B)_{1l}$ for $l \neq i$, and for all l we have $a_{1l}\mathrm{cof}(A)_{1l} = kb_{1l}\mathrm{cof}(B)_{1l}$. By (3.1), we have $\det A = k \det B$.

(3) This is a consequence of (1). If two rows of A are identical, then A is equal to the matrix obtained by interchanging those two rows and therefore by (1) $\det A = -\det A$. This implies $\det A = 0$.

(4) Assume (4) is true for all $n-1 \times n-1$ matrices and fix A and B such that A is obtained by multiplying ith row of B by k and adding it to jth row of B ($i \neq j$) then $\det A = \det B$. If $k = 0$ then $A = B$ and there is nothing to prove, so we may assume $k \neq 0$.

Let C be the matrix obtained by replacing the jth row of B by the ith row of B multiplied by k. By Lemma 3.32, we have that

$$\det A = \det B + \det C$$

and we 'only' need to show that $\det C = 0$. But ith and jth rows of C are proportional. If D is obtained by multiplying the jth row of C by $\frac{1}{k}$ then by (2) we have $\det C = \frac{1}{k} \det D$ (recall that $k \neq 0$!). But ith and jth rows of D are identical, hence by (3) we have $\det D = 0$ and therefore $\det C = 0$. $\square$

Theorem 3.34

Let A and B be two $n \times n$ matrices. Then

$$\det(AB) = \det(A)\det(B)$$

Proof. If A is an elementary matrix of either type, then multiplying by A on the left has the same effect as performing the corresponding elementary row operation. Therefore the equality $\det(AB) = \det A \det B$ in this case follows by Example 3.30 and Theorem 3.33.

If C is the reduced row-echelon form of A then we can write $A = E_1 \cdot E_2 \cdot \dots \cdot E_m \cdot C$ for some elementary matrices $E_1, \dots, E_m$.

Now we consider two cases.

Assume first that $C = I$. Then $A = E_1 \cdot E_2 \cdot \dots \cdot E_m$ and $AB = E_1 \cdot E_2 \cdot \dots \cdot E_m B$. By applying the above equality m times, and then $m-1$ times, we have that

$$\begin{aligned}
\det AB &= \det E_1 \det E_2 \cdot \det E_m \cdot \det B \\
&= \det(E_1 \cdot E_2 \cdot \dots \cdot E_m) \det B \\
&= \det A \det B.
\end{aligned}$$

Now assume $C \neq I$. Since it is in reduced row-echelon form, its last row consists of zeros and by (4) of Example 3.30 the last row of CB consists of zeros. By Lemma 3.31 we have $\det C = \det(CB) = 0$ and therefore

$$\det A = \det(E_1 \cdot E_2 \cdot E_m) \cdot \det(C) = \det(E_1 \cdot E_2 \cdot E_m) \cdot 0 = 0$$

and also

$$\det AB = \det(E_1 \cdot E_2 \cdot E_m) \cdot \det(CB) = \det(E_1 \cdot E_2 \cdot \dots \cdot E_m) 0 = 0$$

hence $\det AB = 0 = \det A \det B$. $\square$

The same 'machine' used in the previous proof will be used again.

Theorem 3.35

Let A be a matrix where A^T is the transpose of A. Then,

$$\det\left(A^T\right) = \det(A)$$

Proof. Note first that the conclusion is true if A is elementary by (5) of Example 3.30.

Let C be the reduced row-echelon form of A. Then we can write $A = E_1 \cdot E_2 \cdot \cdots \cdot E_m C$. Then $A^T = C^T \cdot E_m^T \cdot \cdots \cdot E_2^T \cdot E_1$. By Theorem 3.34 we have

$$\det(A^T) = \det(C^T) \cdot \det(E_m^T) \cdot \cdots \cdot \det(E_2^T) \cdot \det(E_1).$$

By (5) of Example 3.30 we have that $\det E_j = \det E_j^T$ for all j. Also, $\det C$ is either 0 or 1 (depending on whether $C = I$ or not) and in either case $\det C = \det C^T$. Therefore $\det A = \det A^T$. ☐

The above discussions allow us to now prove Theorem 3.9. It is restated below.

Theorem 3.36

Expanding an $n \times n$ matrix along any row or column always gives the same result, which is the determinant.

Proof. We first show that the determinant can be computed along any row. The case $n = 1$ does not apply and thus let $n \geq 2$.

Let Abe an $n \times n$ matrix and fix $j > 1$. We need to prove that

$$\det A = \sum_{i=1}^{n} a_{j,i} \mathrm{cof}(A)_{j,i}.$$

Let us prove the case when $j = 2$.

Let B be the matrix obtained from A by interchanging its 1st and 2nd rows. Then by Theorem 3.33 we have

$$\det A = -\det B.$$

Now we have

$$\det B = \sum_{i=1}^{n} b_{1,i} \mathrm{cof}(B)_{1,i}.$$

Since B is obtained by interchanging the 1st and 2nd rows of A we have that $b_{1,i} = a_{2,i}$ for all i and one can see that $minor(B)_{1,i} = minor(A)_{2,i}$.

Further,

$$\mathrm{cof}(B)_{1,i} = (-1)^{1+i} minor B_{1,i} = -(-1)^{2+i} minor(A)_{2,i} = -\mathrm{cof}(A)_{2,i}$$

hence $\det B = -\sum_{i=1}^{n} a_{2,i}\mathrm{cof}(A)_{2,i}$, and therefore $\det A = -\det B = \sum_{i=1}^{n} a_{2,i}\mathrm{cof}(A)_{2,i}$ as desired.

The case when $j > 2$ is very similar; we still have $minor(B)_{1,i} = minor(A)_{j,i}$ but checking that $\det B = -\sum_{i=1}^{n} a_{j,i} \mathrm{cof}(A)_{j,i}$ is slightly more involved.

Now the cofactor expansion along column j of A is equal to the cofactor expansion along row j of A^T, which is by the above result just proved equal to the cofactor expansion along row 1 of A^T, which is equal to the cofactor expansion along column 1 of A. Thus the cofactor cofactor along any column yields the same result.

Finally, since $\det A = \det A^T$ by Theorem 3.35, we conclude that the cofactor expansion along row 1 of A is equal to the cofactor expansion along row 1 of A^T, which is equal to the cofactor expansion along column 1 of A. Thus the proof is complete. ☐

3.1.5. Finding Determinants using Row Operations

Theorems 3.16, 3.18 and 3.21 illustrate how row operations affect the determinant of a matrix. In this section, we look at two examples where row operations are used to find the determinant of a large matrix. Recall that when working with large matrices, Laplace Expansion is effective but timely, as there are many steps involved. This section provides useful tools for an alternative method. By first applying row operations, we can obtain a simpler matrix to which we apply Laplace Expansion.

While working through questions such as these, it is useful to record your row operations as you go along. Keep this in mind as you read through the next example.

Example 3.37: Finding a Determinant

Find the determinant of the matrix

$$A = \left[\begin{array}{rrrr} 1 & 2 & 3 & 4 \\ 5 & 1 & 2 & 3 \\ 4 & 5 & 4 & 3 \\ 2 & 2 & -4 & 5 \end{array}\right]$$

Solution. We will use the properties of determinants outlined above to find $\det(A)$. First, add -5 times the first row to the second row. Then add -4 times the first row to the third row, and -2 times the first row to the fourth row. This yields the matrix

$$B = \left[\begin{array}{rrrr} 1 & 2 & 3 & 4 \\ 0 & -9 & -13 & -17 \\ 0 & -3 & -8 & -13 \\ 0 & -2 & -10 & -3 \end{array}\right]$$

Notice that the only row operation we have done so far is adding a multiple of a row to another row. Therefore, by Theorem 3.21, $\det(B) = \det(A)$.

At this stage, you could use Laplace Expansion to find $\det(B)$. However, we will continue with row operations to find an even simpler matrix to work with.

Add -3 times the third row to the second row. By Theorem 3.21 this does not change the value of the determinant. Then, multiply the fourth row by -3. This results in the matrix

$$C = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 11 & 22 \\ 0 & -3 & -8 & -13 \\ 0 & 6 & 30 & 9 \end{bmatrix}$$

Here, $\det(C) = -3\det(B)$, which means that $\det(B) = \left(-\frac{1}{3}\right)\det(C)$

Since $\det(A) = \det(B)$, we now have that $\det(A) = \left(-\frac{1}{3}\right)\det(C)$. Again, you could use Laplace Expansion here to find $\det(C)$. However, we will continue with row operations.

Now replace the add 2 times the third row to the fourth row. This does not change the value of the determinant by Theorem 3.21. Finally switch the third and second rows. This causes the determinant to be multiplied by -1. Thus $\det(C) = -\det(D)$ where

$$D = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -3 & -8 & -13 \\ 0 & 0 & 11 & 22 \\ 0 & 0 & 14 & -17 \end{bmatrix}$$

Hence, $\det(A) = \left(-\frac{1}{3}\right)\det(C) = \left(\frac{1}{3}\right)\det(D)$

You could do more row operations or you could note that this can be easily expanded along the first column. Then, expand the resulting 3×3 matrix also along the first column. This results in

$$\det(D) = 1(-3)\begin{vmatrix} 11 & 22 \\ 14 & -17 \end{vmatrix} = 1485$$

and so $\det(A) = \left(\frac{1}{3}\right)(1485) = 495$. $\square$

You can see that by using row operations, we can simplify a matrix to the point where Laplace Expansion involves only a few steps. In Example 3.37, we also could have continued until the matrix was in upper triangular form, and taken the product of the entries on the main diagonal. Whenever computing the determinant, it is useful to consider all the possible methods and tools.

Consider the next example.

Example 3.38: Find the Determinant

Find the determinant of the matrix

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

Solution. Once again, we will simplify the matrix through row operations. Add -1 times the first row to the second row. Next add -2 times the first row to the third and finally take -3 times the first row and add to the fourth row. This yields

$$B = \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -1 & -1 \\ 0 & -3 & -4 & 1 \\ 0 & -10 & -8 & -4 \end{bmatrix}$$

By Theorem 3.21, $\det(A) = \det(B)$.

Remember you can work with the columns also. Take -5 times the fourth column and add to the second column. This yields

$$C = \begin{bmatrix} 1 & -8 & 3 & 2 \\ 0 & 0 & -1 & -1 \\ 0 & -8 & -4 & 1 \\ 0 & 10 & -8 & -4 \end{bmatrix}$$

By Theorem 3.21 $\det(A) = \det(C)$.

Now take -1 times the third row and add to the top row. This gives.

$$D = \begin{bmatrix} 1 & 0 & 7 & 1 \\ 0 & 0 & -1 & -1 \\ 0 & -8 & -4 & 1 \\ 0 & 10 & -8 & -4 \end{bmatrix}$$

which by Theorem 3.21 has the same determinant as A.

Now, we can find $\det(D)$ by expanding along the first column as follows. You can see that there will be only one non zero term.

$$\det(D) = 1 \det \begin{bmatrix} 0 & -1 & -1 \\ -8 & -4 & 1 \\ 10 & -8 & -4 \end{bmatrix} + 0 + 0 + 0$$

Expanding again along the first column, we have

$$\det(D) = 1\left(0 + 8 \det \begin{bmatrix} -1 & -1 \\ -8 & -4 \end{bmatrix} + 10 \det \begin{bmatrix} -1 & -1 \\ -4 & 1 \end{bmatrix}\right) = -82$$

Now since $\det(A) = \det(D)$, it follows that $\det(A) = -82$. ♢

Remember that you can verify these answers by using Laplace Expansion on A. Similarly, if you first compute the determinant using Laplace Expansion, you can use the row operation method to verify.

Exercises

Exercise 3.1.1 *Find the determinants of the following matrices.*

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

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

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

Exercise 3.1.2 *Let* $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 3 \\ -2 & 5 & 1 \end{bmatrix}$. *Find the following.*

1. $minor(A)_{11}$
2. $minor(A)_{21}$
3. $minor(A)_{32}$
4. $cof(A)_{11}$
5. $cof(A)_{21}$
6. $cof(A)_{32}$

Exercise 3.1.3 *Find the determinants of the following matrices.*

1. $\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 2 \\ 0 & 9 & 8 \end{bmatrix}$

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

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

Exercise 3.1.4 *Find the following determinant by expanding along the first row and second column.*

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix}$$

Exercise 3.1.5 *Find the following determinant by expanding along the first column and third row.*

$$\begin{vmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{vmatrix}$$

Exercise 3.1.6 *Find the following determinant by expanding along the second row and first column.*

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix}$$

Exercise 3.1.7 *Compute the determinant by cofactor expansion. Pick the easiest row or column to use.*

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

Exercise 3.1.8 *Find the determinant of the following matrices.*

1. $A = \left[\begin{array}{rr} 1 & -34 \\ 0 & 2 \end{array} \right]$

2. $A = \left[\begin{array}{rrr} 4 & 3 & 14 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{array} \right]$

3. $A = \left[\begin{array}{rrrr} 2 & 3 & 15 & 0 \\ 0 & 4 & 1 & 7 \\ 0 & 0 & -3 & 5 \\ 0 & 0 & 0 & 1 \end{array} \right]$

Exercise 3.1.9 *An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.*

$$\left[\begin{array}{cc} a & b \\ c & d \end{array} \right] \rightarrow \cdots \rightarrow \left[\begin{array}{cc} a & c \\ b & d \end{array} \right]$$

Exercise 3.1.10 *An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.*

$$\left[\begin{array}{cc} a & b \\ c & d \end{array} \right] \rightarrow \cdots \rightarrow \left[\begin{array}{cc} c & d \\ a & b \end{array} \right]$$

Exercise 3.1.11 *An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.*

$$\left[\begin{array}{cc} a & b \\ c & d \end{array} \right] \rightarrow \cdots \rightarrow \left[\begin{array}{cc} a & b \\ a+c & b+d \end{array} \right]$$

Exercise 3.1.12 *An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.*

$$\left[\begin{array}{cc} a & b \\ c & d \end{array} \right] \rightarrow \cdots \rightarrow \left[\begin{array}{cc} a & b \\ 2c & 2d \end{array} \right]$$

Exercise 3.1.13 *An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.*

$$\left[\begin{array}{cc} a & b \\ c & d \end{array}\right] \rightarrow \cdots \rightarrow \left[\begin{array}{cc} b & a \\ d & c \end{array}\right]$$

Exercise 3.1.14 *Let A be an $r \times r$ matrix and suppose there are $r-1$ rows (columns) such that all rows (columns) are linear combinations of these $r-1$ rows (columns). Show* $\det(A) = 0$.

Exercise 3.1.15 *Show* $\det(aA) = a^n \det(A)$ *for an $n \times n$ matrix A and scalar a.*

Exercise 3.1.16 *Construct 2×2 matrices A and B to show that the* $\det A \det B = \det(AB)$.

Exercise 3.1.17 *Is it true that* $\det(A+B) = \det(A) + \det(B)$? *If this is so, explain why. If it is not so, give a counter example.*

Exercise 3.1.18 *An $n \times n$ matrix is called* ***nilpotent*** *if for some positive integer, k it follows $A^k = 0$. If A is a nilpotent matrix and k is the smallest possible integer such that $A^k = 0$, what are the possible values of* $\det(A)$?

Exercise 3.1.19 *A matrix is said to be* ***orthogonal*** *if $A^T A = I$. Thus the inverse of an orthogonal matrix is just its transpose. What are the possible values of* $\det(A)$ *if A is an orthogonal matrix?*

Exercise 3.1.20 *Let A and B be two $n \times n$ matrices. $A \sim B$ (A is* ***similar*** *to B) means there exists an invertible matrix P such that $A = P^{-1}BP$. Show that if $A \sim B$, then* $\det(A) = \det(B)$.

Exercise 3.1.21 *Tell whether each statement is true or false. If true, provide a proof. If false, provide a counter example.*

1. *If A is a 3×3 matrix with a zero determinant, then one column must be a multiple of some other column.*

2. *If any two columns of a square matrix are equal, then the determinant of the matrix equals zero.*

3. *For two $n \times n$ matrices A and B,* $\det(A+B) = \det(A) + \det(B)$.

4. *For an $n \times n$ matrix A,* $\det(3A) = 3\det(A)$

5. *If A^{-1} exists then* $\det(A^{-1}) = \det(A)^{-1}$.

6. *If B is obtained by multiplying a single row of A by 4 then* $\det(B) = 4\det(A)$.

7. *For A an $n \times n$ matrix,* $\det(-A) = (-1)^n \det(A)$.

8. *If A is a real $n \times n$ matrix, then* $\det\left(A^T A\right) \geq 0$.

9. *If* $A^k = 0$ *for some positive integer* k, *then* $\det(A) = 0$.
10. *If* $AX = 0$ *for some* $X \neq 0$, *then* $\det(A) = 0$.

Exercise 3.1.22 *Find the determinant using row operations to first simplify.*

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

Exercise 3.1.23 *Find the determinant using row operations to first simplify.*

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

Exercise 3.1.24 *Find the determinant using row operations to first simplify.*

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

Exercise 3.1.25 *Find the determinant using row operations to first simplify.*

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

3.2 Applications of the Determinant

Outcomes

A. *Use determinants to determine whether a matrix has an inverse, and evaluate the inverse using cofactors.*

B. *Apply Cramer's Rule to solve a* 2×2 *or a* 3×3 *linear system.*

C. *Given data points, find an appropriate interpolating polynomial and use it to estimate points.*

3.2.1. A Formula for the Inverse

The determinant of a matrix also provides a way to find the inverse of a matrix. Recall the definition of the inverse of a matrix in Definition 2.33. We say that A^{-1}, an $n \times n$ matrix, is the inverse of A, also $n \times n$, if $AA^{-1} = I$ and $A^{-1}A = I$.

We now define a new matrix called the **cofactor matrix** of A. The cofactor matrix of A is the matrix whose ij^{th} entry is the ij^{th} cofactor of A. The formal definition is as follows.

Definition 3.39: The Cofactor Matrix

*Let $A = [a_{ij}]$ be an $n \times n$ matrix. Then the **cofactor matrix of** A, denoted $cof(A)$, is defined by $cof(A) = \left[cof(A)_{ij} \right]$ where $cof(A)_{ij}$ is the ij^{th} cofactor of A.*

Note that $\operatorname{cof}(A)_{ij}$ denotes the ij^{th} entry of the cofactor matrix.

We will use the cofactor matrix to create a formula for the inverse of A. First, we define the **adjugate** of A to be the transpose of the cofactor matrix. We can also call this matrix the **classical adjoint** of A, and we denote it by $\operatorname{adj}(A)$.

In the specific case where A is a 2×2 matrix given by

$$A = \left[\begin{array}{cc} a & b \\ c & d \end{array} \right]$$

then $\operatorname{adj}(A)$ is given by

$$\operatorname{adj}(A) = \left[\begin{array}{rr} d & -b \\ -c & a \end{array} \right]$$

In general, $\operatorname{adj}(A)$ can always be found by taking the transpose of the cofactor matrix of A. The following theorem provides a formula for A^{-1} using the determinant and adjugate of A.

Theorem 3.40: The Inverse and the Determinant

Let A be an $n \times n$ matrix. Then

$$A \; adj(A) = adj(A) \, A = \det(A) I$$

Moreover A is invertible if and only if $\det(A) \neq 0$. In this case we have:

$$A^{-1} = \frac{1}{\det(A)} adj(A)$$

Notice that the first formula holds for any $n \times n$ matrix A, and in the case A is invertible we actually have a formula for A^{-1}.

Consider the following example.

Example 3.41: Find Inverse Using the Determinant

Find the inverse of the matrix

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

using the formula in Theorem 3.40.

Solution. According to Theorem 3.40,

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A)$$

First we will find the determinant of this matrix. Using Theorems 3.16, 3.18, and 3.21, we can first simplify the matrix through row operations. First, add -3 times the first row to the second row. Then add -1 times the first row to the third row to obtain

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -6 & -8 \\ 0 & 0 & -2 \end{bmatrix}$$

By Theorem 3.21, $\det(A) = \det(B)$. By Theorem 3.13, $\det(B) = 1 \times -6 \times -2 = 12$. Hence, $\det(A) = 12$.

Now, we need to find $\operatorname{adj}(A)$. To do so, first we will find the cofactor matrix of A. This is given by

$$\operatorname{cof}(A) = \begin{bmatrix} -2 & -2 & 6 \\ 4 & -2 & 0 \\ 2 & 8 & -6 \end{bmatrix}$$

Here, the ij^{th} entry is the ij^{th} cofactor of the original matrix A which you can verify. Therefore, from Theorem 3.40, the inverse of A is given by

$$A^{-1} = \frac{1}{12} \begin{bmatrix} -2 & -2 & 6 \\ 4 & -2 & 0 \\ 2 & 8 & -6 \end{bmatrix}^T = \begin{bmatrix} -\frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & \frac{2}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

Remember that we can always verify our answer for A^{-1}. Compute the product AA^{-1} and $A^{-1}A$ and make sure each product is equal to I.

Compute $A^{-1}A$ as follows

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

You can verify that $AA^{-1} = I$ and hence our answer is correct. $\square$

We will look at another example of how to use this formula to find A^{-1}.

Example 3.42: Find the Inverse From a Formula

Find the inverse of the matrix

$$A = \left[\begin{array}{rrr} \frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{6} & \frac{1}{3} & -\frac{1}{2} \\ -\frac{5}{6} & \frac{2}{3} & -\frac{1}{2} \end{array} \right]$$

using the formula given in Theorem 3.40.

Solution. First we need to find $\det(A)$. This step is left as an exercise and you should verify that $\det(A) = \frac{1}{6}$. The inverse is therefore equal to

$$A^{-1} = \frac{1}{(1/6)} \operatorname{adj}(A) = 6 \operatorname{adj}(A)$$

We continue to calculate as follows. Here we show the 2×2 determinants needed to find the cofactors.

$$A^{-1} = 6 \left[\begin{array}{rrr} \left| \begin{array}{rr} \frac{1}{3} & -\frac{1}{2} \\ \frac{2}{3} & -\frac{1}{2} \end{array} \right| & -\left| \begin{array}{rr} -\frac{1}{6} & -\frac{1}{2} \\ -\frac{5}{6} & -\frac{1}{2} \end{array} \right| & \left| \begin{array}{rr} -\frac{1}{6} & \frac{1}{3} \\ -\frac{5}{6} & \frac{2}{3} \end{array} \right| \\ -\left| \begin{array}{rr} 0 & \frac{1}{2} \\ \frac{2}{3} & -\frac{1}{2} \end{array} \right| & \left| \begin{array}{rr} \frac{1}{2} & \frac{1}{2} \\ -\frac{5}{6} & -\frac{1}{2} \end{array} \right| & -\left| \begin{array}{rr} \frac{1}{2} & 0 \\ -\frac{5}{6} & \frac{2}{3} \end{array} \right| \\ \left| \begin{array}{rr} 0 & \frac{1}{2} \\ \frac{1}{3} & -\frac{1}{2} \end{array} \right| & -\left| \begin{array}{rr} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{6} & -\frac{1}{2} \end{array} \right| & \left| \begin{array}{rr} \frac{1}{2} & 0 \\ -\frac{1}{6} & \frac{1}{3} \end{array} \right| \end{array} \right]^T$$

Expanding all the 2×2 determinants, this yields

$$A^{-1} = 6 \left[\begin{array}{rrr} \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{6} & -\frac{1}{3} \\ -\frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{array} \right]^T = \left[\begin{array}{rrr} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -2 & 1 \end{array} \right]$$

Again, you can always check your work by multiplying $A^{-1}A$ and AA^{-1} and ensuring these products equal I.

$$A^{-1}A = \left[\begin{array}{rrr} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -2 & 1 \end{array} \right] \left[\begin{array}{rrr} \frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{6} & \frac{1}{3} & -\frac{1}{2} \\ -\frac{5}{6} & \frac{2}{3} & -\frac{1}{2} \end{array} \right] = \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right]$$

This tells us that our calculation for A^{-1} is correct. It is left to the reader to verify that $AA^{-1} = I$. ◻

The verification step is very important, as it is a simple way to check your work! If you multiply $A^{-1}A$ and AA^{-1} and these products are not both equal to I, be sure to go back and double check each step. One common error is to forget to take the transpose of the cofactor matrix, so be sure to complete this step.

We will now prove Theorem 3.40.

Proof. (of Theorem 3.40) Recall that the (i, j)-entry of $\operatorname{adj}(A)$ is equal to $\operatorname{cof}(A)_{ji}$. Thus the (i, j)-entry of $B = A \cdot \operatorname{adj}(A)$ is :

$$B_{ij} = \sum_{k=1}^{n} a_{ik} \operatorname{adj}(A)_{kj} = \sum_{k=1}^{n} a_{ik} \operatorname{cof}(A)_{jk}$$

By the cofactor expansion theorem, we see that this expression for B_{ij} is equal to the determinant of the matrix obtained from A by replacing its jth row by $a_{i1}, a_{i2}, \ldots a_{in}$ — i.e., its ith row.

If $i = j$ then this matrix is A itself and therefore $B_{ii} = \det A$. If on the other hand $i \neq j$, then this matrix has its ith row equal to its jth row, and therefore $B_{ij} = 0$ in his case. Thus we obtain:

$$A \operatorname{adj}(A) = \det(A) I$$

Similarly we can verify that:

$$\operatorname{adj}(A) A = \det(A) I$$

And this proves the first part of the theorem.

Further if A is invertible, then by Theorem 3.24 we have:

$$1 = \det(I) = \det\left(AA^{-1}\right) = \det(A) \det\left(A^{-1}\right)$$

and thus $\det(A) \neq 0$. Equivalently, if $\det(A) = 0$, then A is not invertible.

Finally if $\det(A) \neq 0$, then the above formula shows that A is invertible and that:

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A)$$

This completes the proof. ◻

This method for finding the inverse of A is useful in many contexts. In particular, it is useful with complicated matrices where the entries are functions, rather than numbers.

Consider the following example.

Example 3.43: Inverse for Non-Constant Matrix

Suppose

$$A(t) = \left[\begin{array}{ccc} e^t & 0 & 0 \\ 0 & \cos t & \sin t \\ 0 & -\sin t & \cos t \end{array}\right]$$

Show that $A(t)^{-1}$ exists and then find it.

Solution. First note $\det(A(t)) = e^t(\cos^2 t + \sin^2 t) = e^t \neq 0$ so $A(t)^{-1}$ exists.

The cofactor matrix is

$$C(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{bmatrix}$$

and so the inverse is

$$\frac{1}{e^t} \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{bmatrix}^T = \begin{bmatrix} e^{-t} & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{bmatrix}$$

□

3.2.2. Cramer's Rule

Another context in which the formula given in Theorem 3.40 is important is **Cramer's Rule**. Recall that we can represent a system of linear equations in the form $AX = B$, where the solutions to this system are given by X. Cramer's Rule gives a formula for the solutions X in the special case that A is a square invertible matrix. Note this rule does not apply if you have a system of equations in which there is a different number of equations than variables (in other words, when A is not square), or when A is not invertible.

Suppose we have a system of equations given by $AX = B$, and we want to find solutions X which satisfy this system. Then recall that if A^{-1} exists,

$$\begin{aligned} AX &= B \\ A^{-1}(AX) &= A^{-1}B \\ \left(A^{-1}A\right)X &= A^{-1}B \\ IX &= A^{-1}B \\ X &= A^{-1}B \end{aligned}$$

Hence, the solutions X to the system are given by $X = A^{-1}B$. Since we assume that A^{-1} exists, we can use the formula for A^{-1} given above. Substituting this formula into the equation for X, we have

$$X = A^{-1}B = \frac{1}{\det(A)} \operatorname{adj}(A) B$$

Let x_i be the i^{th} entry of X and b_j be the j^{th} entry of B. Then this equation becomes

$$x_i = \sum_{j=1}^{n} \left[a_{ij}\right]^{-1} b_j = \sum_{j=1}^{n} \frac{1}{\det(A)} \operatorname{adj}(A)_{ij}\, b_j$$

where $\operatorname{adj}(A)_{ij}$ is the ij^{th} entry of $\operatorname{adj}(A)$.

By the formula for the expansion of a determinant along a column,

$$x_i = \frac{1}{\det(A)} \det \begin{bmatrix} * & \cdots & b_1 & \cdots & * \\ \vdots & & \vdots & & \vdots \\ * & \cdots & b_n & \cdots & * \end{bmatrix}$$

where here the i^{th} column of A is replaced with the column vector $[b_1 \cdots \cdot, b_n]^T$. The determinant of this modified matrix is taken and divided by $\det(A)$. This formula is known as Cramer's rule.

We formally define this method now.

Procedure 3.44: Using Cramer's Rule

Suppose A is an $n \times n$ invertible matrix and we wish to solve the system $AX = B$ for $X = [x_1, \cdots, x_n]^T$. Then Cramer's rule says

$$x_i = \frac{\det(A_i)}{\det(A)}$$

where A_i is the matrix obtained by replacing the i^{th} column of A with the column matrix

$$B = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

We illustrate this procedure in the following example.

Example 3.45: Using Cramer's Rule

Find x, y, z if

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

Solution. We will use method outlined in Procedure 3.44 to find the values for x, y, z which give the solution to this system. Let

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

In order to find x, we calculate

$$x = \frac{\det(A_1)}{\det(A)}$$

where A_1 is the matrix obtained from replacing the first column of A with B.

Hence, A_1 is given by

$$A_1 = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 2 & 1 \\ 3 & -3 & 2 \end{bmatrix}$$

Therefore,

$$x = \frac{\det(A_1)}{\det(A)} = \frac{\begin{vmatrix} 1 & 2 & 1 \\ 2 & 2 & 1 \\ 3 & -3 & 2 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{vmatrix}} = \frac{1}{2}$$

Similarly, to find y we construct A_2 by replacing the second column of A with B. Hence, A_2 is given by

$$A_2 = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 2 & 1 \\ 2 & 3 & 2 \end{bmatrix}$$

Therefore,

$$y = \frac{\det(A_2)}{\det(A)} = \frac{\begin{vmatrix} 1 & 1 & 1 \\ 3 & 2 & 1 \\ 2 & 3 & 2 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{vmatrix}} = -\frac{1}{7}$$

Similarly, A_3 is constructed by replacing the third column of A with B. Then, A_3 is given by

$$A_3 = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 2 & 2 \\ 2 & -3 & 3 \end{bmatrix}$$

Therefore, z is calculated as follows.

$$z = \frac{\det(A_3)}{\det(A)} = \frac{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 2 \\ 2 & -3 & 3 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{vmatrix}} = \frac{11}{14}$$

$\square$

Cramer's Rule gives you another tool to consider when solving a system of linear equations.

We can also use Cramer's Rule for systems of non linear equations. Consider the following system where the matrix A has functions rather than numbers for entries.

Example 3.46: Use Cramer's Rule for Non-Constant Matrix

Solve for z if

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix}$$

Solution. We are asked to find the value of z in the solution. We will solve using Cramer's rule. Thus

$$z = \frac{\begin{vmatrix} 1 & 0 & 1 \\ 0 & e^t \cos t & t \\ 0 & -e^t \sin t & t^2 \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{vmatrix}} = t\left((\cos t)\, t + \sin t\right) e^{-t}$$

□

Exercises

Exercise 3.2.1 *Let*

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

Determine whether the matrix A has an inverse by finding whether the determinant is non zero. If the determinant is nonzero, find the inverse using the formula for the inverse which involves the cofactor matrix.

Exercise 3.2.2 *Let*

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

Determine whether the matrix A has an inverse by finding whether the determinant is non zero. If the determinant is nonzero, find the inverse using the formula for the inverse.

Exercise 3.2.3 *Let*

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

Determine whether the matrix A has an inverse by finding whether the determinant is non zero. If the determinant is nonzero, find the inverse using the formula for the inverse.

Exercise 3.2.4 *Let*

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

Determine whether the matrix A has an inverse by finding whether the determinant is non zero. If the determinant is nonzero, find the inverse using the formula for the inverse.

Exercise 3.2.5 *Let*

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

Determine whether the matrix A has an inverse by finding whether the determinant is non zero. If the determinant is nonzero, find the inverse using the formula for the inverse.

Exercise 3.2.6 *For the following matrices, determine if they are invertible. If so, use the formula for the inverse in terms of the cofactor matrix to find each inverse. If the inverse does not exist, explain why.*

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

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

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

Exercise 3.2.7 *Consider the matrix*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{bmatrix}$$

Does there exist a value of t for which this matrix fails to have an inverse? Explain.

Exercise 3.2.8 *Consider the matrix*

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

Does there exist a value of t for which this matrix fails to have an inverse? Explain.

Exercise 3.2.9 *Consider the matrix*

$$A = \begin{bmatrix} e^t & \cosh t & \sinh t \\ e^t & \sinh t & \cosh t \\ e^t & \cosh t & \sinh t \end{bmatrix}$$

Does there exist a value of t for which this matrix fails to have an inverse? Explain.

Exercise 3.2.10 *Consider the matrix*

$$A = \begin{bmatrix} e^t & e^{-t}\cos t & e^{-t}\sin t \\ e^t & -e^{-t}\cos t - e^{-t}\sin t & -e^{-t}\sin t + e^{-t}\cos t \\ e^t & 2e^{-t}\sin t & -2e^{-t}\cos t \end{bmatrix}$$

Does there exist a value of t for which this matrix fails to have an inverse? Explain.

Exercise 3.2.11 *Show that if $\det(A) \neq 0$ for A an $n \times n$ matrix, it follows that if $AX = 0$, then $X = 0$.*

Exercise 3.2.12 *Suppose A, B are $n \times n$ matrices and that $AB = I$. Show that then $BA = I$.* ***Hint:*** *First explain why $\det(A), \det(B)$ are both nonzero. Then $(AB)A = A$ and then show $BA(BA - I) = 0$. From this use what is given to conclude $A(BA - I) = 0$. Then use Problem 3.2.11.*

Exercise 3.2.13 *Use the formula for the inverse in terms of the cofactor matrix to find the inverse of the matrix*

$$A = \begin{bmatrix} e^t & 0 & 0 \\ 0 & e^t\cos t & e^t\sin t \\ 0 & e^t\cos t - e^t\sin t & e^t\cos t + e^t\sin t \end{bmatrix}$$

Exercise 3.2.14 *Find the inverse, if it exists, of the matrix*

$$A = \begin{bmatrix} e^t & \cos t & \sin t \\ e^t & -\sin t & \cos t \\ e^t & -\cos t & -\sin t \end{bmatrix}$$

Exercise 3.2.15 *Suppose A is an upper triangular matrix. Show that A^{-1} exists if and only if all elements of the main diagonal are non zero. Is it true that A^{-1} will also be upper triangular? Explain. Could the same be concluded for lower triangular matrices?*

Exercise 3.2.16 *If A, B, and C are each $n \times n$ matrices and ABC is invertible, show why each of A, B, and C are invertible.*

Exercise 3.2.17 *Decide if this statement is true or false: Cramer's rule is useful for finding solutions to systems of linear equations in which there is an infinite set of solutions.*

Exercise 3.2.18 *Use Cramer's rule to find the solution to*

$$\begin{gathered} x + 2y = 1 \\ 2x - y = 2 \end{gathered}$$

Exercise 3.2.19 *Use Cramer's rule to find the solution to*

$$\begin{gathered} x + 2y + z = 1 \\ 2x - y - z = 2 \\ x + z = 1 \end{gathered}$$

4. $\mathbb{R}^n$

4.1 Vectors in $\mathbb{R}^n$

Outcomes

A. *Find the position vector of a point in $\mathbb{R}^n$.*

The notation $\mathbb{R}^n$ refers to the collection of ordered lists of n real numbers, that is

$$\mathbb{R}^n = \{(x_1 \cdots x_n) : x_j \in \mathbb{R} \text{ for } j = 1, \cdots, n\}$$

In this chapter, we take a closer look at vectors in $\mathbb{R}^n$. First, we will consider what $\mathbb{R}^n$ looks like in more detail. Recall that the point given by $0 = (0, \cdots, 0)$ is called the **origin**.

Now, consider the case of $\mathbb{R}^n$ for $n = 1$. Then from the definition we can identify $\mathbb{R}$ with points in $\mathbb{R}^1$ as follows:

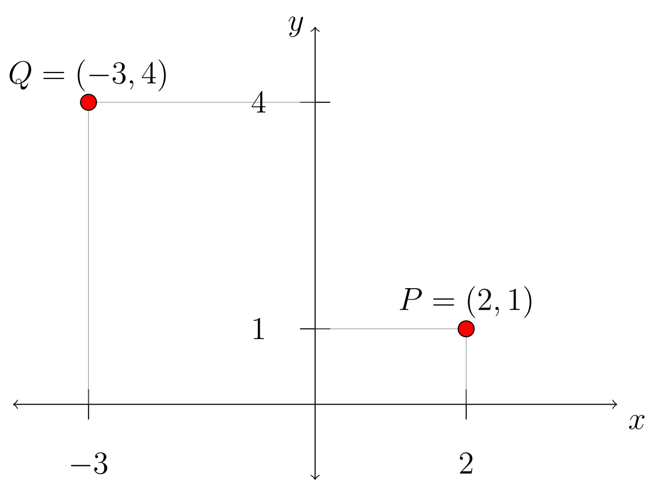
$$\mathbb{R} = \mathbb{R}^1 = \{(x_1) : x_1 \in \mathbb{R}\}$$

Hence, $\mathbb{R}$ is defined as the set of all real numbers and geometrically, we can describe this as all the points on a line.

Now suppose $n = 2$. Then, from the definition,

$$\mathbb{R}^2 = \{(x_1, x_2) : x_j \in \mathbb{R} \text{ for } j = 1, 2\}$$

Consider the familiar coordinate plane, with an x axis and a y axis. Any point within this coordinate plane is identified by where it is located along the x axis, and also where it is located along the y axis. Consider as an example the following diagram.



Hence, every element in $\mathbb{R}^2$ is identified by two components, x and y, in the usual manner. The coordinates x, y (or x_1,x_2) uniquely determine a point in the plan. Note that while the definition uses x_1 and x_2 to label the coordinates and you may be used to x and y, these notations are equivalent.

Now suppose $n = 3$. You may have previously encountered the 3-dimensional coordinate system, given by

$$\mathbb{R}^3 = \{(x_1, x_2, x_3) : x_j \in \mathbb{R} \text{ for } j = 1, 2, 3\}$$

Points in $\mathbb{R}^3$ will be determined by three coordinates, often written (x, y, z) which correspond to the x, y, and z axes. We can think as above that the first two coordinates determine a point in a plane. The third component determines the height above or below the plane, depending on whether this number is positive or negative, and all together this determines a point in space. You see that the ordered triples correspond to points in space just as the ordered pairs correspond to points in a plane and single real numbers correspond to points on a line.

The idea behind the more general $\mathbb{R}^n$ is that we can extend these ideas beyond $n = 3$. This discussion regarding points in $\mathbb{R}^n$ leads into a study of vectors in $\mathbb{R}^n$. While we consider $\mathbb{R}^n$ for all n, we will largely focus on $n = 2, 3$ in this section.

Consider the following definition.

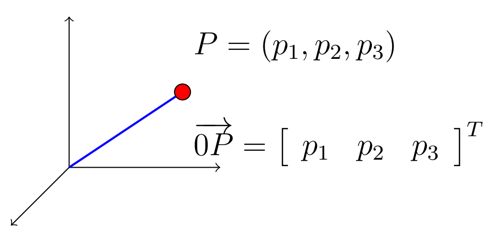
Definition 4.1: The Position Vector

Let $P = (p_1, \cdots, p_n)$ be the coordinates of a point in $\mathbb{R}^n$. Then the vector $\overrightarrow{0P}$ with its tail at $0 = (0, \cdots, 0)$ and its tip at P is called the ***position vector*** *of the point P. We write*

$$\overrightarrow{0P} = \left[\begin{array}{c} p_1 \\ \vdots \\ p_n \end{array} \right]$$

For this reason we may write both $P = (p_1, \cdots, p_n) \in \mathbb{R}^n$ and $\overrightarrow{0P} = [p_1 \cdots p_n]^T \in \mathbb{R}^n$.

This definition is illustrated in the following picture for the special case of $\mathbb{R}^3$.

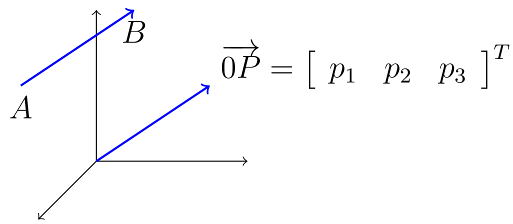


Thus every point P in $\mathbb{R}^n$ determines its position vector $\overrightarrow{0P}$. Conversely, every such position vector $\overrightarrow{0P}$ which has its tail at 0 and point at P determines the point P of $\mathbb{R}^n$.

Now suppose we are given two points, P, Q whose coordinates are $(p_1, \cdots, p_n)$ and $(q_1, \cdots, q_n)$ respectively. We can also determine the **position vector from P to Q** (also called the **vector from P to Q**) defined as follows.

$$\overrightarrow{PQ} = \left[\begin{array}{c} q_1 - p_1 \\ \vdots \\ q_n - p_n \end{array} \right] = \overrightarrow{0Q} - \overrightarrow{0P}$$

Now, imagine taking a vector in $\mathbb{R}^n$ and moving it around, always keeping it pointing in the same direction as shown in the following picture.



After moving it around, it is regarded as the same vector. Each vector, $\overrightarrow{0P}$ and $\overrightarrow{AB}$ has the same length (or magnitude) and direction. Therefore, they are equal.

Consider now the general definition for a vector in $\mathbb{R}^n$.

Definition 4.2: Vectors in $\mathbb{R}^n$

Let $\mathbb{R}^n = \left\{ (x_1, \cdots, x_n) : x_j \in \mathbb{R} \text{ for } j = 1, \cdots, n \right\}$. Then,

$$\vec{x} = \left[\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array} \right]$$

*is called a **vector**. Vectors have both size (magnitude) and direction. The numbers x_j are called the **components** of $\vec{x}$.*

Using this notation, we may use $\vec{p}$ to denote the position vector of point P. Notice that in this context, $\vec{p} = \overrightarrow{0P}$. These notations may be used interchangeably.

You can think of the components of a vector as directions for obtaining the vector. Consider $n = 3$. Draw a vector with its tail at the point $(0,0,0)$ and its tip at the point (a,b,c). This vector it is obtained by starting at $(0,0,0)$, moving parallel to the x axis to $(a,0,0)$ and then from here, moving parallel to the y axis to $(a,b,0)$ and finally parallel to the z axis to (a,b,c). Observe that the same vector would result if you began at the point (d,e,f), moved parallel to the x axis to $(d+a,e,f)$, then parallel to the y axis to $(d+a,e+b,f)$, and finally parallel to the z axis to $(d+a,e+b,f+c)$. Here, the vector would have its tail sitting at the point determined by $A = (d,e,f)$ and its point at $B = (d+a,e+b,f+c)$. It is the **same vector** because it will point in the same direction and have the same length. It is like you took an actual arrow, and moved it from one location to another keeping it pointing the same direction.

We conclude this section with a brief discussion regarding notation. In previous sections, we have written vectors as columns, or $n \times 1$ matrices. For convenience in this chapter we may write vectors as the transpose of row vectors, or $1 \times n$ matrices. These are of course equivalent and we may move between both notations. Therefore, recognize that

$$\left[\begin{array}{c} 2 \\ 3 \end{array} \right] = \left[\begin{array}{cc} 2 & 3 \end{array} \right]^T$$

Notice that two vectors $\vec{u} = [u_1 \cdots u_n]^T$ and $\vec{v} = [v_1 \cdots v_n]^T$ are equal if and only if all corresponding components are equal. Precisely,

$$\begin{array}{c} \vec{u} = \vec{v} \text{ if and only if} \\ u_j = v_j \text{ for all } j = 1, \cdots, n \end{array}$$

Thus $\left[\begin{array}{ccc} 1 & 2 & 4 \end{array} \right]^T \in \mathbb{R}^3$ and $\left[\begin{array}{ccc} 2 & 1 & 4 \end{array} \right]^T \in \mathbb{R}^3$ but $\left[\begin{array}{ccc} 1 & 2 & 4 \end{array} \right]^T \neq \left[\begin{array}{ccc} 2 & 1 & 4 \end{array} \right]^T$ because, even though the same numbers are involved, the order of the numbers is different.

For the specific case of $\mathbb{R}^3$, there are three special vectors which we often use. They are given by

$$\begin{aligned} \vec{i} &= \left[\begin{array}{ccc} 1 & 0 & 0 \end{array} \right]^T \\ \vec{j} &= \left[\begin{array}{ccc} 0 & 1 & 0 \end{array} \right]^T \\ \vec{k} &= \left[\begin{array}{ccc} 0 & 0 & 1 \end{array} \right]^T \end{aligned}$$

We can write any vector $\vec{u} = \left[\begin{array}{ccc} u_1 & u_2 & u_3 \end{array} \right]^T$ as a linear combination of these vectors, written as $\vec{u} = u_1\vec{i} + u_2\vec{j} + u_3\vec{k}$. This notation will be used throughout this chapter.

4.2 Algebra in $\mathbb{R}^n$

Outcomes

A. *Understand vector addition and scalar multiplication, algebraically.*

B. *Introduce the notion of linear combination of vectors.*

Addition and scalar multiplication are two important algebraic operations done with vectors. Notice that these operations apply to vectors in $\mathbb{R}^n$, for any value of n. We will explore these operations in more detail in the following sections.

4.2.1. Addition of Vectors in $\mathbb{R}^n$

Addition of vectors in $\mathbb{R}^n$ is defined as follows.

Definition 4.3: Addition of Vectors in $\mathbb{R}^n$

If $\vec{u} = \left[\begin{array}{c} u_1 \\ \vdots \\ u_n \end{array}\right]$, $\vec{v} = \left[\begin{array}{c} v_1 \\ \vdots \\ v_n \end{array}\right] \in \mathbb{R}^n$ *then* $\vec{u}+\vec{v} \in \mathbb{R}^n$ *and is defined by*

$$\begin{aligned} \vec{u}+\vec{v} &= \left[\begin{array}{c} u_1 \\ \vdots \\ u_n \end{array}\right] + \left[\begin{array}{c} v_1 \\ \vdots \\ v_n \end{array}\right] \\ &= \left[\begin{array}{c} u_1+v_1 \\ \vdots \\ u_n+v_n \end{array}\right] \end{aligned}$$

To add vectors, we simply add corresponding components exactly as we did for matrices. Therefore, in order to add vectors, they must be the same size.

Similarly to matrices, addition of vectors satisfies some important properties. These are outlined in the following theorem.

Theorem 4.4: Properties of Vector Addition

The following properties hold for vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$.

- *The Commutative Law of Addition*

$$\vec{u}+\vec{v} = \vec{v}+\vec{u}$$

- *The Associative Law of Addition*

$$(\vec{u}+\vec{v})+\vec{w} = \vec{u}+(\vec{v}+\vec{w})$$

- *The Existence of an Additive Identity*

$$\vec{u}+\vec{0} = \vec{u} \tag{4.1}$$

- *The Existence of an Additive Inverse*

$$\vec{u}+(-\vec{u}) = \vec{0}$$

The proof of this theorem follows from the similar properties for matrix operations. Thus the additive identity shown in equation 4.1 is also called the **zero vector**, the $n \times 1$ vector in which all components are equal to 0. Further, $-\vec{u}$ is simply the vector with all components having same value as those of $\vec{u}$ but opposite sign; this is just $(-1)\vec{u}$. This will be made more explicit in the next section when we explore scalar multiplication of vectors. Note that subtraction is defined as $\vec{u}-\vec{v} = \vec{u}+(-\vec{v})$.

4.2.2. Scalar Multiplication of Vectors in $\mathbb{R}^n$

Scalar multiplication of vectors in $\mathbb{R}^n$ is defined as follows. Notice that, just like addition, this definition is the same as the corresponding definition for matrices.

Definition 4.5: Scalar Multiplication of Vectors in $\mathbb{R}^n$

If $\vec{u} \in \mathbb{R}^n$ and $k \in \mathbb{R}$ is a scalar, then $k\vec{u} \in \mathbb{R}^n$ is defined by

$$k\vec{u} = k \left[\begin{array}{c} u_1 \\ \vdots \\ u_n \end{array} \right] = \left[\begin{array}{c} ku_1 \\ \vdots \\ ku_n \end{array} \right]$$

Just as with addition, scalar multiplication of vectors satisfies several important properties. These are outlined in the following theorem.

Theorem 4.6: Properties of Scalar Multiplication

The following properties hold for vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ and k, p scalars.

- *The Distributive Law over Vector Addition*

$$k\left(\vec{u} + \vec{v}\right) = k\vec{u} + k\vec{v}$$

- *The Distributive Law over Scalar Addition*

$$\left(k + p\right)\vec{u} = k\vec{u} + p\vec{u}$$

- *The Associative Law for Scalar Multiplication*

$$k\left(p\vec{u}\right) = \left(kp\right)\vec{u}$$

- *Rule for Multiplication by 1*

$$1\vec{u} = \vec{u}$$

Proof: Again the verification of these properties follows from the corresponding properties for scalar multiplication of matrices.

As a refresher we can show that

$$k\left(\vec{u} + \vec{v}\right) = k\vec{u} + k\vec{v}$$

Note that:

$$\begin{aligned} k\left(\vec{u} + \vec{v}\right) &= k\left[u_1 + v_1 \cdots u_n + v_n\right]^T \\ &= \left[k\left(u_1 + v_1\right) \cdots k\left(u_n + v_n\right)\right]^T \\ &= \left[ku_1 + kv_1 \cdots ku_n + kv_n\right]^T \\ &= \left[ku_1 \cdots ku_n\right]^T + \left[kv_1 \cdots kv_n\right]^T \\ &= k\vec{u} + k\vec{v} \end{aligned}$$

$\blacksquare$

We now present a useful notion you may have seen earlier combining vector addition and scalar multiplication

Definition 4.7: Linear Combination

A vector $\vec{v}$ is said to be a ***linear combination*** *of the vectors $\vec{u}_1, \cdots, \vec{u}_n$ if there exist scalars, $a_1, \cdots, a_n$ such that*

$$\vec{v} = a_1\vec{u}_1 + \cdots + a_n\vec{u}_n$$

For example,

$$3\begin{bmatrix} -4 \\ 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -18 \\ 3 \\ 2 \end{bmatrix}.$$

Thus we can say that

$$\vec{v} = \begin{bmatrix} -18 \\ 3 \\ 2 \end{bmatrix}$$

is a linear combination of the vectors

$$\vec{u}_1 = \begin{bmatrix} -4 \\ 1 \\ 0 \end{bmatrix} \text{ and } \vec{u}_2 = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$

Exercises

Exercise 4.2.1 *Find* $-3\begin{bmatrix} 5 \\ -1 \\ 2 \\ -3 \end{bmatrix} + 5\begin{bmatrix} -8 \\ 2 \\ -3 \\ 6 \end{bmatrix}$.

Exercise 4.2.2 *Find* $-7\begin{bmatrix} 6 \\ 0 \\ 4 \\ -1 \end{bmatrix} + 6\begin{bmatrix} -13 \\ -1 \\ 1 \\ 6 \end{bmatrix}$.

Exercise 4.2.3 *Decide whether*

$$\vec{v} = \begin{bmatrix} 4 \\ 4 \\ -3 \end{bmatrix}$$

is a linear combination of the vectors

$$\vec{u}_1 = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix} \text{ and } \vec{u}_2 = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}.$$

Exercise 4.2.4 *Decide whether*

$$\vec{v} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}$$

is a linear combination of the vectors

$$\vec{u}_1 = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix} \text{ and } \vec{u}_2 = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}.$$

4.3 Geometric Meaning of Vector Addition

Outcomes

A. *Understand vector addition, geometrically.*

Recall that an element of $\mathbb{R}^n$ is an ordered list of numbers. For the specific case of $n = 2, 3$ this can be used to determine a point in two or three dimensional space. This point is specified relative to some coordinate axes.

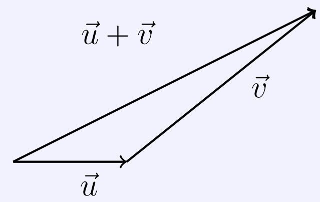
Consider the case $n = 3$. Recall that taking a vector and moving it around without changing its length or direction does not change the vector. This is important in the geometric representation of vector addition.

Suppose we have two vectors, $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^3$. Each of these can be drawn geometrically by placing the tail of each vector at 0 and its point at (u_1, u_2, u_3) and (v_1, v_2, v_3) respectively. Suppose we slide the vector $\vec{v}$ so that its tail sits at the point of $\vec{u}$. We know that this does not change the vector $\vec{v}$. Now, draw a new vector from the tail of $\vec{u}$ to the point of $\vec{v}$. This vector is $\vec{u} + \vec{v}$.

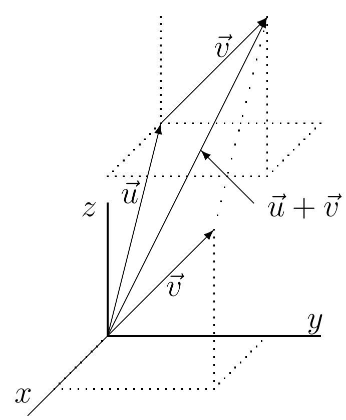
The geometric significance of vector addition in $\mathbb{R}^n$ for any n is given in the following definition.

Definition 4.8: Geometry of Vector Addition

Let $\vec{u}$ and $\vec{v}$ be two vectors. Slide $\vec{v}$ so that the tail of $\vec{v}$ is on the point of $\vec{u}$. Then draw the arrow which goes from the tail of $\vec{u}$ to the point of $\vec{v}$. This arrow represents the vector $\vec{u}+\vec{v}$.



This definition is illustrated in the following picture in which $\vec{u}+\vec{v}$ is shown for the special case $n = 3$.

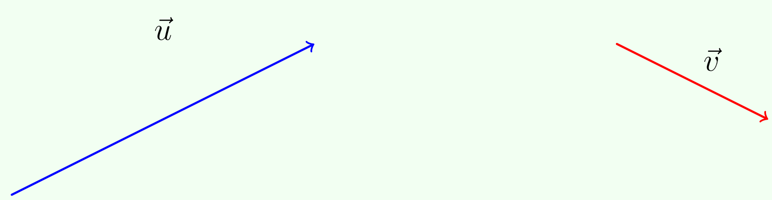


Notice the parallelogram created by $\vec{u}$ and $\vec{v}$ in the above diagram. Then $\vec{u}+\vec{v}$ is the directed diagonal of the parallelogram determined by the two vectors $\vec{u}$ and $\vec{v}$.

When you have a vector $\vec{v}$, its additive inverse $-\vec{v}$ will be the vector which has the same magnitude as $\vec{v}$ but the opposite direction. When one writes $\vec{u}-\vec{v}$, the meaning is $\vec{u}+(-\vec{v})$ as with real numbers. The following example illustrates these definitions and conventions.

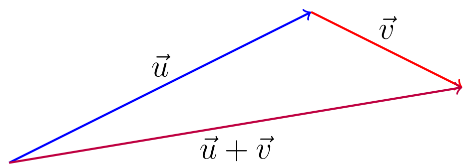
Example 4.9: Graphing Vector Addition

Consider the following picture of vectors $\vec{u}$ and $\vec{v}$.

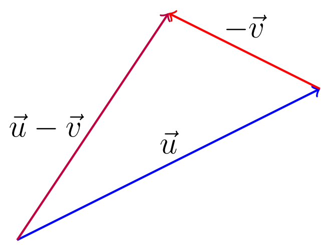


Sketch a picture of $\vec{u}+\vec{v}, \vec{u}-\vec{v}$.

Solution. We will first sketch $\vec{u}+\vec{v}$. Begin by drawing $\vec{u}$ and then at the point of $\vec{u}$, place the tail of $\vec{v}$ as shown. Then $\vec{u}+\vec{v}$ is the vector which results from drawing a vector from the tail of $\vec{u}$ to the tip of $\vec{v}$.



Next consider $\vec{u}-\vec{v}$. This means $\vec{u}+(-\vec{v})$. From the above geometric description of vector addition, $-\vec{v}$ is the vector which has the same length but which points in the opposite direction to $\vec{v}$. Here is a picture.



$\square$

4.4 Length of a Vector

Outcomes

A. *Find the length of a vector and the distance between two points in $\mathbb{R}^n$.*

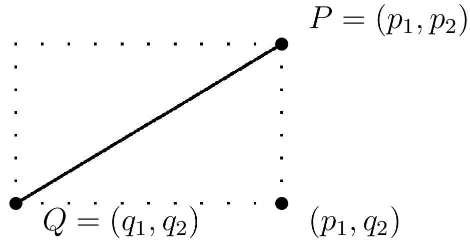
B. *Find the corresponding unit vector to a vector in $\mathbb{R}^n$.*

In this section, we explore what is meant by the length of a vector in $\mathbb{R}^n$. We develop this concept by first looking at the distance between two points in $\mathbb{R}^n$.

First, we will consider the concept of distance for $\mathbb{R}$, that is, for points in $\mathbb{R}^1$. Here, the distance between two points P and Q is given by the absolute value of their difference. We denote the distance between P and Q by $d(P,Q)$ which is defined as

$$d(P,Q)=\sqrt{(P-Q)^2} \tag{4.2}$$

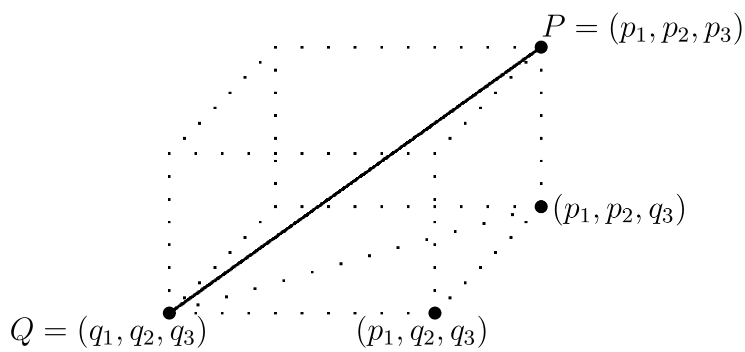
Consider now the case for $n=2$, demonstrated by the following picture.



There are two points $P = (p_1, p_2)$ and $Q = (q_1, q_2)$ in the plane. The distance between these points is shown in the picture as a solid line. Notice that this line is the hypotenuse of a right triangle which is half of the rectangle shown in dotted lines. We want to find the length of this hypotenuse which will give the distance between the two points. Note the lengths of the sides of this triangle are $|p_1 - q_1|$ and $|p_2 - q_2|$, the absolute value of the difference in these values. Therefore, the Pythagorean Theorem implies the length of the hypotenuse (and thus the distance between P and Q) equals

$$\left(|p_1 - q_1|^2 + |p_2 - q_2|^2\right)^{1/2} = \left((p_1 - q_1)^2 + (p_2 - q_2)^2\right)^{1/2} \tag{4.3}$$

Now suppose $n = 3$ and let $P = (p_1, p_2, p_3)$ and $Q = (q_1, q_2, q_3)$ be two points in $\mathbb{R}^3$. Consider the following picture in which the solid line joins the two points and a dotted line joins the points (q_1, q_2, q_3) and (p_1, p_2, q_3).



Here, we need to use Pythagorean Theorem twice in order to find the length of the solid line. First, by the Pythagorean Theorem, the length of the dotted line joining (q_1, q_2, q_3) and (p_1, p_2, q_3) equals

$$\left((p_1 - q_1)^2 + (p_2 - q_2)^2\right)^{1/2}$$

while the length of the line joining (p_1, p_2, q_3) to (p_1, p_2, p_3) is just $|p_3 - q_3|$. Therefore, by the Pythagorean Theorem again, the length of the line joining the points $P = (p_1, p_2, p_3)$ and $Q = (q_1, q_2, q_3)$ equals

$$\left(\left(\left((p_1 - q_1)^2 + (p_2 - q_2)^2\right)^{1/2}\right)^2 + (p_3 - q_3)^2\right)^{1/2}$$

$$= \left((p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2\right)^{1/2} \tag{4.4}$$

This discussion motivates the following definition for the distance between points in $\mathbb{R}^n$.

Definition 4.10: Distance Between Points

Let $P = (p_1, \cdots, p_n)$ and $Q = (q_1, \cdots, q_n)$ be two points in $\mathbb{R}^n$. Then the distance between these points is defined as

$$\text{distance between } P \text{ and } Q = d(P,Q) = \left(\sum_{k=1}^{n} |p_k - q_k|^2 \right)^{1/2}$$

This is called the ***distance formula****. We may also write $|P - Q|$ as the distance between P and Q.*

From the above discussion, you can see that Definition 4.10 holds for the special cases $n = 1, 2, 3$, as in Equations 4.2, 4.3, 4.4. In the following example, we use Definition 4.10 to find the distance between two points in $\mathbb{R}^4$.

Example 4.11: Distance Between Points

Find the distance between the points P and Q in $\mathbb{R}^4$, where P and Q are given by

$$P = (1, 2, -4, 6)$$

and

$$Q = (2, 3, -1, 0)$$

Solution. We will use the formula given in Definition 4.10 to find the distance between P and Q. Use the distance formula and write

$$d(P,Q) = \left((1-2)^2 + (2-3)^2 + (-4-(-1))^2 + (6-0)^2 \right)^{\frac{1}{2}} = 47$$

Therefore, $d(P,Q) = \sqrt{47}$.

$\square$

There are certain properties of the distance between points which are important in our study. These are outlined in the following theorem.

Theorem 4.12: Properties of Distance

Let P and Q be points in $\mathbb{R}^n$, and let the distance between them, $d(P,Q)$, be given as in Definition 4.10. Then, the following properties hold .

- $d(P,Q) = d(Q,P)$
- $d(P,Q) \geq 0$, *and equals 0 exactly when $P = Q$.*

There are many applications of the concept of distance. For instance, given two points, we can ask what collection of points are all the same distance between the given points. This is explored in the following example.

Example 4.13: The Plane Between Two Points

Describe the points in $\mathbb{R}^3$ *which are at the same distance between* $(1,2,3)$ *and* $(0,1,2)$.

Solution. Let $P=(p_1,p_2,p_3)$ be such a point. Therefore, P is the same distance from $(1,2,3)$ and $(0,1,2)$. Then by Definition 4.10,

$$\sqrt{(p_1-1)^2+(p_2-2)^2+(p_3-3)^2}=\sqrt{(p_1-0)^2+(p_2-1)^2+(p_3-2)^2}$$

Squaring both sides we obtain

$$(p_1-1)^2+(p_2-2)^2+(p_3-3)^2=p_1^2+(p_2-1)^2+(p_3-2)^2$$

and so

$$p_1^2-2p_1+14+p_2^2-4p_2+p_3^2-6p_3=p_1^2+p_2^2-2p_2+5+p_3^2-4p_3$$

Simplifying, this becomes

$$-2p_1+14-4p_2-6p_3=-2p_2+5-4p_3$$

which can be written as

$$2p_1+2p_2+2p_3=-9 \tag{4.5}$$

Therefore, the points $P=(p_1,p_2,p_3)$ which are the same distance from each of the given points form a plane whose equation is given by 4.5. ◻

We can now use our understanding of the distance between two points to define what is meant by the length of a vector. Consider the following definition.

Definition 4.14: Length of a Vector

Let $\vec{u}=[u_1\cdots u_n]^T$ *be a vector in* $\mathbb{R}^n$. *Then, the length of* $\vec{u}$, *written* $\|\vec{u}\|$ *is given by*

$$\|\vec{u}\|=\sqrt{u_1^2+\cdots+u_n^2}$$

This definition corresponds to Definition 4.10, if you consider the vector $\vec{u}$ to have its tail at the point $0=(0,\cdots,0)$ and its tip at the point $U=(u_1,\cdots,u_n)$. Then the length of $\vec{u}$ is equal to the distance between 0 and U, $d(0,U)$. In general, $d(P,Q)=\|\overrightarrow{PQ}\|$.

Consider Example 4.11. By Definition 4.14, we could also find the distance between P and Q as the length of the vector connecting them. Hence, if we were to draw a vector $\overrightarrow{PQ}$ with its tail at P and its point at Q, this vector would have length equal to $\sqrt{47}$.

We conclude this section with a new definition for the special case of vectors of length 1.

Definition 4.15: Unit Vector

Let $\vec{u}$ *be a vector in* $\mathbb{R}^n$. *Then, we call* $\vec{u}$ *a* ***unit vector*** *if it has length 1, that is if*

$$\|\vec{u}\|=1$$

Let $\vec{v}$ be a vector in $\mathbb{R}^n$. Then, the vector $\vec{u}$ which has the same direction as $\vec{v}$ but length equal to 1 is the corresponding unit vector of $\vec{v}$. This vector is given by

$$\vec{u} = \frac{1}{\|\vec{v}\|}\vec{v}$$

We often use the term **normalize** to refer to this process. When we **normalize** a vector, we find the corresponding unit vector of length 1. Consider the following example.

Example 4.16: Finding a Unit Vector

Let $\vec{v}$ be given by

$$\vec{v} = \left[\begin{array}{ccc} 1 & -3 & 4 \end{array}\right]^T$$

Find the unit vector $\vec{u}$ which has the same direction as $\vec{v}$.

Solution. We will use Definition 4.15 to solve this. Therefore, we need to find the length of $\vec{v}$ which, by Definition 4.14 is given by

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2}$$

Using the corresponding values we find that

$$\begin{array}{rcl} \|\vec{v}\| & = & \sqrt{1^2 + (-3)^2 + 4^2} \\ & = & \sqrt{1+9+16} \\ & = & \sqrt{26} \end{array}$$

In order to find $\vec{u}$, we divide $\vec{v}$ by $\sqrt{26}$. The result is

$$\begin{array}{rcl} \vec{u} & = & \frac{1}{\|\vec{v}\|}\vec{v} \\ & = & \frac{1}{\sqrt{26}}\left[\begin{array}{ccc} 1 & -3 & 4 \end{array}\right]^T \\ & = & \left[\begin{array}{ccc} \frac{1}{\sqrt{26}} & -\frac{3}{\sqrt{26}} & \frac{4}{\sqrt{26}} \end{array}\right]^T \end{array}$$

You can verify using the Definition 4.14 that $\|\vec{u}\| = 1$. ∎

4.5 Geometric Meaning of Scalar Multiplication

Outcomes

A. *Understand scalar multiplication, geometrically.*

Recall that the point $P = (p_1, p_2, p_3)$ determines a vector $\vec{p}$ from 0 to P. The length of $\vec{p}$, denoted $\|\vec{p}\|$, is equal to $\sqrt{p_1^2 + p_2^2 + p_3^2}$ by Definition 4.10.

Now suppose we have a vector $\vec{u} = \left[\begin{array}{ccc} u_1 & u_2 & u_3 \end{array} \right]^T$ and we multiply $\vec{u}$ by a scalar k. By Definition 4.5, $k\vec{u} = \left[\begin{array}{ccc} ku_1 & ku_2 & ku_3 \end{array} \right]^T$. Then, by using Definition 4.10, the length of this vector is given by

$$\sqrt{\left((ku_1)^2 + (ku_2)^2 + (ku_3)^2 \right)} = |k| \sqrt{u_1^2 + u_2^2 + u_3^2}$$

Thus the following holds.

$$\|k\vec{u}\| = |k| \, \|\vec{u}\|$$

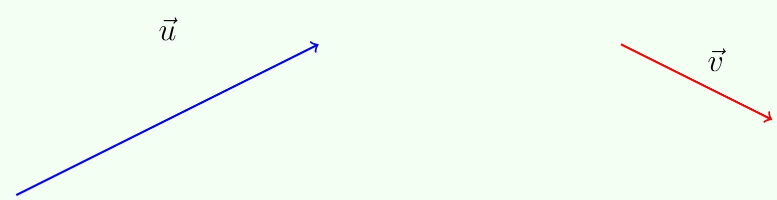
In other words, multiplication by a scalar magnifies or shrinks the length of the vector by a factor of $|k|$. If $|k| > 1$, the length of the resulting vector will be magnified. If $|k| < 1$, the length of the resulting vector will shrink. Remember that by the definition of the absolute value, $|k| > 0$.

What about the direction? Draw a picture of $\vec{u}$ and $k\vec{u}$ where k is negative. Notice that this causes the resulting vector to point in the opposite direction while if $k > 0$ it preserves the direction the vector points. Therefore the direction can either reverse, if $k < 0$, or remain preserved, if $k > 0$.

Consider the following example.

Example 4.17: Graphing Scalar Multiplication

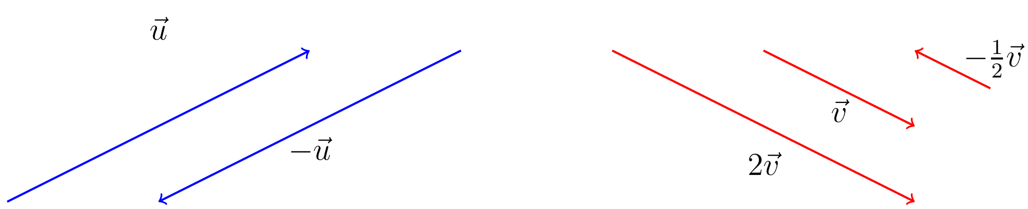
Consider the vectors $\vec{u}$ and $\vec{v}$ drawn below.



Draw $-\vec{u}$, $2\vec{v}$, and $-\frac{1}{2}\vec{v}$.

Solution.

In order to find $-\vec{u}$, we preserve the length of $\vec{u}$ and simply reverse the direction. For $2\vec{v}$, we double the length of $\vec{v}$, while preserving the direction. Finally $-\frac{1}{2}\vec{v}$ is found by taking half the length of $\vec{v}$ and reversing the direction. These vectors are shown in the following diagram.



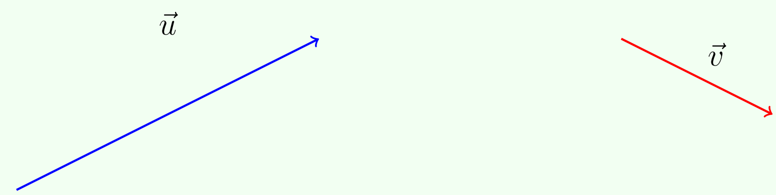
□

Now that we have studied both vector addition and scalar multiplication, we can combine the two actions. Recall Definition 4.7 of linear combinations of column matrices. We can apply this definition to vectors in $\mathbb{R}^n$. A linear combination of vectors in $\mathbb{R}^n$ is a sum of vectors multiplied by scalars.

In the following example, we examine the geometric meaning of this concept.

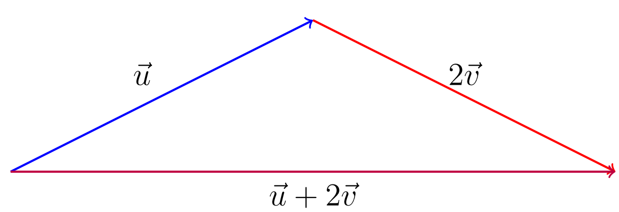
Example 4.18: Graphing a Linear Combination of Vectors

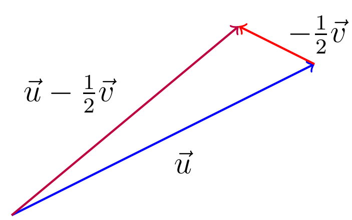
Consider the following picture of the vectors $\vec{u}$ and $\vec{v}$



Sketch a picture of $\vec{u}+2\vec{v}, \vec{u}-\frac{1}{2}\vec{v}$.

Solution. The two vectors are shown below.





□

4.6 Parametric Lines

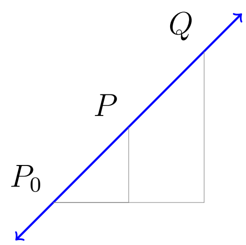
Outcomes

A. *Find the vector and parametric equations of a line.*

We can use the concept of vectors and points to find equations for arbitrary lines in $\mathbb{R}^n$, although in this section the focus will be on lines in $\mathbb{R}^3$.

To begin, consider the case $n = 1$ so we have $\mathbb{R}^1 = \mathbb{R}$. There is only one line here which is the familiar number line, that is $\mathbb{R}$ itself. Therefore it is not necessary to explore the case of $n = 1$ further.

Now consider the case where $n = 2$, in other words $\mathbb{R}^2$. Let P and P_0 be two different points in $\mathbb{R}^2$ which are contained in a line L. Let $\vec{p}$ and $\vec{p_0}$ be the position vectors for the points P and P_0 respectively. Suppose that Q is an arbitrary point on L. Consider the following diagram.



Our goal is to be able to define Q in terms of P and P_0. Consider the vector $\overrightarrow{P_0P} = \vec{p} - \vec{p_0}$ which has its tail at P_0 and point at P. If we add $\vec{p} - \vec{p_0}$ to the position vector $\vec{p_0}$ for P_0, the sum would be a vector with its point at P. In other words,

$$\vec{p} = \vec{p_0} + (\vec{p} - \vec{p_0})$$

Now suppose we were to add $t(\vec{p} - \vec{p_0})$ to $\vec{p}$ where t is some scalar. You can see that by doing so, we could find a vector with its point at Q. In other words, we can find t such that

$$\vec{q} = \vec{p_0} + t\,(\vec{p} - \vec{p_0})$$

This equation determines the line L in $\mathbb{R}^2$. In fact, it determines a line L in $\mathbb{R}^n$. Consider the following definition.

Definition 4.19: Vector Equation of a Line

Suppose a line L in $\mathbb{R}^n$ contains the two different points P and P_0. Let $\vec{p}$ and $\vec{p_0}$ be the position vectors of these two points, respectively. Then, L is the collection of points Q which have the position vector $\vec{q}$ given by

$$\vec{q} = \vec{p_0} + t\,(\vec{p} - \vec{p_0})$$

where $t \in \mathbb{R}$.
Let $\vec{d} = \vec{p} - \vec{p_0}$. Then $\vec{d}$ is the ***direction vector for*** *L and the* ***vector equation*** *for L is given by*

$$\vec{p} = \vec{p_0} + t\vec{d}, t \in \mathbb{R}$$

Note that this definition agrees with the usual notion of a line in two dimensions and so this is consistent with earlier concepts. Consider now points in $\mathbb{R}^3$. If a point $P \in \mathbb{R}^3$ is given by $P = (x, y, z)$, $P_0 \in \mathbb{R}^3$ by $P_0 = (x_0, y_0, z_0)$, then we can write

$$\left[\begin{array}{c} x \\ y \\ z \end{array}\right] = \left[\begin{array}{c} x_0 \\ y_0 \\ z_0 \end{array}\right] + t \left[\begin{array}{c} a \\ b \\ c \end{array}\right]$$

where $\vec{d} = \left[\begin{array}{c} a \\ b \\ c \end{array}\right]$. This is the vector equation of L written in **component form** .

The following theorem claims that such an equation is in fact a line.

Proposition 4.20: Algebraic Description of a Straight Line

Let $\vec{a}, \vec{b} \in \mathbb{R}^n$ with $\vec{b} \neq \vec{0}$. Then $\vec{x} = \vec{a} + t\vec{b}$, $t \in \mathbb{R}$, is a line.

Proof. Let $\vec{x_1}, \vec{x_2} \in \mathbb{R}^n$. Define $\vec{x_1} = \vec{a}$ and let $\vec{x_2} - \vec{x_1} = \vec{b}$. Since $\vec{b} \neq \vec{0}$, it follows that $\vec{x_2} \neq \vec{x_1}$. Then $\vec{a} + t\vec{b} = \vec{x_1} + t(\vec{x_2} - \vec{x_1})$. It follows that $\vec{x} = \vec{a} + t\vec{b}$ is a line containing the two different points X_1 and X_2 whose position vectors are given by $\vec{x}_1$ and $\vec{x}_2$ respectively. $\square$

We can use the above discussion to find the equation of a line when given two distinct points. Consider the following example.

Example 4.21: A Line From Two Points

Find a vector equation for the line through the points $P_0 = (1, 2, 0)$ and $P = (2, -4, 6)$.

Solution. We will use the definition of a line given above in Definition 4.19 to write this line in the form

$$\vec{q} = \vec{p_0} + t(\vec{p} - \vec{p_0})$$

Let $\vec{q} = \left[\begin{array}{c} x \\ y \\ z \end{array}\right]$. Then, we can find $\vec{p}$ and $\vec{p_0}$ by taking the position vectors of points P and P_0 respectively. Then,

$$\vec{q} = \vec{p_0} + t(\vec{p} - \vec{p_0})$$

can be written as

$$\left[\begin{array}{c} x \\ y \\ z \end{array}\right] = \left[\begin{array}{c} 1 \\ 2 \\ 0 \end{array}\right] + t \left[\begin{array}{r} 1 \\ -6 \\ 6 \end{array}\right], \ t \in \mathbb{R}$$

Here, the direction vector $\left[\begin{array}{r} 1 \\ -6 \\ 6 \end{array}\right]$ is obtained by $\vec{p} - \vec{p_0} = \left[\begin{array}{r} 2 \\ -4 \\ 6 \end{array}\right] - \left[\begin{array}{c} 1 \\ 2 \\ 0 \end{array}\right]$ as indicated above in Definition 4.19. $\square$

Notice that in the above example we said that we found "a" vector equation for the line, not "the" equation. The reason for this terminology is that there are infinitely many different vector equations for the same line. To see this, replace t with another parameter, say $3s$. Then you obtain a different vector equation for the same line because the same set of points is obtained.

In Example 4.21, the vector given by $\left[\begin{array}{r} 1 \\ -6 \\ 6 \end{array}\right]$ is the direction vector defined in Definition 4.19. If we know the direction vector of a line, as well as a point on the line, we can find the vector equation.

Consider the following example.

Example 4.22: A Line From a Point and a Direction Vector

Find a vector equation for the line which contains the point $P_0 = (1,2,0)$ *and has direction vector* $\vec{d} = \left[\begin{array}{c} 1 \\ 2 \\ 1 \end{array}\right]$

Solution. We will use Definition 4.19 to write this line in the form $\vec{p} = \vec{p_0} + t\vec{d},\ t \in \mathbb{R}$. We are given the direction vector $\vec{d}$. In order to find $\vec{p_0}$, we can use the position vector of the point P_0. This is given by $\left[\begin{array}{c} 1 \\ 2 \\ 0 \end{array}\right]$. Letting $\vec{p} = \left[\begin{array}{c} x \\ y \\ z \end{array}\right]$, the equation for the line is given by

$$\left[\begin{array}{c} x \\ y \\ z \end{array}\right] = \left[\begin{array}{c} 1 \\ 2 \\ 0 \end{array}\right] + t\left[\begin{array}{c} 1 \\ 2 \\ 1 \end{array}\right],\ t \in \mathbb{R} \tag{4.6}$$

□

We sometimes elect to write a line such as the one given in 4.6 in the form

$$\left.\begin{array}{l} x = 1 + t \\ y = 2 + 2t \\ z = t \end{array}\right\} \text{ where } t \in \mathbb{R} \tag{4.7}$$

This set of equations give the same information as 4.6, and is called the **parametric equation of the line**.

Consider the following definition.

Definition 4.23: Parametric Equation of a Line

Let L be a line in $\mathbb{R}^3$ which has direction vector $\vec{d} = \left[\begin{array}{c} a \\ b \\ c \end{array}\right]$ and goes through the point $P_0 = (x_0, y_0, z_0)$. Then, letting t be a parameter, we can write L as

$$\left.\begin{array}{l} x = x_0 + ta \\ y = y_0 + tb \\ z = z_0 + tc \end{array}\right\} \text{ where } t \in \mathbb{R}$$

*This is called a **parametric equation** of the line L.*

You can verify that the form discussed following Example 4.22 in equation 4.7 is of the form given in Definition 4.23.

There is one other form for a line which is useful, which is the **symmetric form**. Consider the line given by 4.7. You can solve for the parameter t to write

$$\begin{array}{l} t = x - 1 \\ t = \frac{y-2}{2} \\ t = z \end{array}$$

Therefore,

$$x - 1 = \frac{y-2}{2} = z$$

This is the **symmetric form**of the line.

In the following example, we look at how to take the equation of a line from symmetric form to parametric form.

Example 4.24: Change Symmetric Form to Parametric Form

*Suppose the **symmetric form of a line** is*

$$\frac{x-2}{3} = \frac{y-1}{2} = z + 3$$

Write the line in parametric form as well as vector form.

Solution. We want to write this line in the form given by Definition 4.23. This is of the form

$$\left.\begin{array}{l} x = x_0 + ta \\ y = y_0 + tb \\ z = z_0 + tc \end{array}\right\} \text{ where } t \in \mathbb{R}$$

Let $t = \frac{x-2}{3}, t = \frac{y-1}{2}$ and $t = z + 3$, as given in the symmetric form of the line. Then solving for x, y, z, yields

$$\left.\begin{array}{l} x = 2 + 3t \\ y = 1 + 2t \\ z = -3 + t \end{array}\right\} \text{ with } t \in \mathbb{R}$$

This is the parametric equation for this line.

Now, we want to write this line in the form given by Definition 4.19. This is the form

$$\vec{p} = \vec{p_0} + t\vec{d}$$

where $t \in \mathbb{R}$. This equation becomes

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} + t \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix},\ t \in \mathbb{R}$$

□

Exercises

Exercise 4.6.1 *Find the vector equation for the line through* $(-7,6,0)$ *and* $(-1,1,4)$. *Then, find the parametric equations for this line.*

Exercise 4.6.2 *Find parametric equations for the line through the point* $(7,7,1)$ *with a direction vector* $\vec{d} = \begin{bmatrix} 1 \\ 6 \\ 2 \end{bmatrix}$.

Exercise 4.6.3 *Parametric equations of the line are*

$$\begin{aligned} x &= t+2 \\ y &= 6-3t \\ z &= -t-6 \end{aligned}$$

Find a direction vector for the line and a point on the line.

Exercise 4.6.4 *Find the vector equation for the line through the two points* $(-5,5,1)$, $(2,2,4)$. *Then, find the parametric equations.*

Exercise 4.6.5 *The equation of a line in two dimensions is written as* $y = x - 5$. *Find parametric equations for this line.*

Exercise 4.6.6 *Find parametric equations for the line through* $(6,5,-2)$ *and* $(5,1,2)$.

Exercise 4.6.7 *Find the vector equation and parametric equations for the line through the point* $(-7,10,-6)$ *with a direction vector* $\vec{d} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$.

Exercise 4.6.8 *Parametric equations of the line are*

$$\begin{aligned} x &= 2t+2 \\ y &= 5-4t \\ z &= -t-3 \end{aligned}$$

Find a direction vector for the line and a point on the line, and write the vector equation of the line.

Exercise 4.6.9 *Find the vector equation and parametric equations for the line through the two points* $(4,10,0)$, $(1,-5,-6)$.

Exercise 4.6.10 *Find the point on the line segment from* $P=(-4,7,5)$ *to* $Q=(2,-2,-3)$ *which is* $\frac{1}{7}$ *of the way from* P *to* Q.

Exercise 4.6.11 *Suppose a triangle in* $\mathbb{R}^n$ *has vertices at* P_1, P_2, *and* P_3. *Consider the lines which are drawn from a vertex to the mid point of the opposite side. Show these three lines intersect in a point and find the coordinates of this point.*

4.7 The Dot Product

Outcomes

A. *Compute the dot product of vectors, and use this to compute vector projections.*

4.7.1. The Dot Product

There are two ways of multiplying vectors which are of great importance in applications. The first of these is called the **dot product**. When we take the dot product of vectors, the result is a scalar. For this reason, the dot product is also called the **scalar product** and sometimes the **inner product**. The definition is as follows.

Definition 4.25: Dot Product

Let $\vec{u},\vec{v}$ *be two vectors in* $\mathbb{R}^n$. *Then we define the* ***dot product*** $\vec{u}\bullet\vec{v}$ *as*

$$\vec{u}\bullet\vec{v} = \sum_{k=1}^{n} u_k v_k$$

The dot product $\vec{u}\bullet\vec{v}$ is sometimes denoted as $(\vec{u},\vec{v})$ where a comma replaces $\bullet$. It can also be written as $\langle\vec{u},\vec{v}\rangle$. If we write the vectors as column or row matrices, it is equal to the matrix product $\vec{v}\vec{w}^T$.

Consider the following example.

Example 4.26: Compute a Dot Product

Find $\vec{u} \bullet \vec{v}$ *for*

$$\vec{u} = \left[\begin{array}{r} 1 \\ 2 \\ 0 \\ -1 \end{array}\right], \vec{v} = \left[\begin{array}{r} 0 \\ 1 \\ 2 \\ 3 \end{array}\right]$$

Solution. By Definition 4.25, we must compute

$$\vec{u} \bullet \vec{v} = \sum_{k=1}^{4} u_k v_k$$

This is given by

$$\begin{aligned} \vec{u} \bullet \vec{v} &= (1)(0) + (2)(1) + (0)(2) + (-1)(3) \\ &= 0 + 2 + 0 + -3 \\ &= -1 \end{aligned}$$

□

With this definition, there are several important properties satisfied by the dot product.

Proposition 4.27: Properties of the Dot Product

Let k *and* p *denote scalars and* $\vec{u}, \vec{v}, \vec{w}$ *denote vectors. Then the dot product* $\vec{u} \bullet \vec{v}$ *satisfies the following properties.*

- $\vec{u} \bullet \vec{v} = \vec{v} \bullet \vec{u}$
- $\vec{u} \bullet \vec{u} \geq 0$ *and equals zero if and only if* $\vec{u} = \vec{0}$
- $(k\vec{u} + p\vec{v}) \bullet \vec{w} = k(\vec{u} \bullet \vec{w}) + p(\vec{v} \bullet \vec{w})$
- $\vec{u} \bullet (k\vec{v} + p\vec{w}) = k(\vec{u} \bullet \vec{v}) + p(\vec{u} \bullet \vec{w})$
- $\|\vec{u}\|^2 = \vec{u} \bullet \vec{u}$

The proof is left as an exercise. This proposition tells us that we can also use the dot product to find the length of a vector.

Example 4.28: Length of a Vector

Find the length of

$$\vec{u} = \left[\begin{array}{c} 2 \\ 1 \\ 4 \\ 2 \end{array} \right]$$

That is, find $\|\vec{u}\|$.

Solution. By Proposition 4.27, $\|\vec{u}\|^2 = \vec{u} \bullet \vec{u}$. Therefore, $\|\vec{u}\| = \sqrt{\vec{u} \bullet \vec{u}}$. First, compute $\vec{u} \bullet \vec{u}$.

This is given by

$$\begin{aligned} \vec{u} \bullet \vec{u} &= (2)(2) + (1)(1) + (4)(4) + (2)(2) \\ &= 4 + 1 + 16 + 4 \\ &= 25 \end{aligned}$$

Then,

$$\begin{aligned} \|\vec{u}\| &= \sqrt{\vec{u} \bullet \vec{u}} \\ &= \sqrt{25} \\ &= 5 \end{aligned}$$

□

You may wish to compare this to our previous definition of length, given in Definition 4.14.

The **Cauchy Schwarz inequality** is a fundamental inequality satisfied by the dot product. It is given in the following theorem.

Theorem 4.29: Cauchy Schwarz Inequality

The dot product satisfies the inequality

$$|\vec{u} \bullet \vec{v}| \leq \|\vec{u}\| \|\vec{v}\| \tag{4.8}$$

Furthermore equality is obtained if and only if one of $\vec{u}$ *or* $\vec{v}$ *is a scalar multiple of the other.*

Proof. First note that if $\vec{v} = \vec{0}$ both sides of 4.8 equal zero and so the inequality holds in this case. Therefore, it will be assumed in what follows that $\vec{v} \neq \vec{0}$.

Define a function of $t \in \mathbb{R}$ by

$$f(t) = (\vec{u} + t\vec{v}) \bullet (\vec{u} + t\vec{v})$$

Then by Proposition 4.27, $f(t) \geq 0$ for all $t \in \mathbb{R}$. Also from Proposition 4.27

$$\begin{aligned} f(t) &= \vec{u} \bullet (\vec{u} + t\vec{v}) + t\vec{v} \bullet (\vec{u} + t\vec{v}) \\ &= \vec{u} \bullet \vec{u} + t(\vec{u} \bullet \vec{v}) + t\vec{v} \bullet \vec{u} + t^2 \vec{v} \bullet \vec{v} \\ &= \|\vec{u}\|^2 + 2t(\vec{u} \bullet \vec{v}) + \|\vec{v}\|^2 t^2 \end{aligned}$$

Now this means the graph of $y = f(t)$ is a parabola which opens up and either its vertex touches the t axis or else the entire graph is above the t axis. In the first case, there exists some t where $f(t) = 0$ and this requires $\vec{u} + t\vec{v} = \vec{0}$ so one vector is a multiple of the other. Then clearly equality holds in 4.8. In the case where $\vec{v}$ is not a multiple of $\vec{u}$, it follows $f(t) > 0$ for all t which says $f(t)$ has no real zeros and so from the quadratic formula,

$$(2(\vec{u} \bullet \vec{v}))^2 - 4\|\vec{u}\|^2\|\vec{v}\|^2 < 0$$

which is equivalent to $|\vec{u} \bullet \vec{v}| < \|\vec{u}\|\|\vec{v}\|$. $\square$

Notice that this proof was based only on the properties of the dot product listed in Proposition 4.27. This means that whenever an operation satisfies these properties, the Cauchy Schwarz inequality holds. There are many other instances of these properties besides vectors in $\mathbb{R}^n$.

The Cauchy Schwarz inequality provides another proof of the **triangle inequality** for distances in $\mathbb{R}^n$.

Theorem 4.30: Triangle Inequality

For $\vec{u}, \vec{v} \in \mathbb{R}^n$

$$\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\| \tag{4.9}$$

and equality holds if and only if one of the vectors is a non-negative scalar multiple of the other.
Also

$$\left|\|\vec{u}\| - \|\vec{v}\|\right| \leq \|\vec{u} - \vec{v}\| \tag{4.10}$$

Proof. By properties of the dot product and the Cauchy Schwarz inequality,

$$\begin{array}{rcl}
\|\vec{u} + \vec{v}\|^2 & = & (\vec{u} + \vec{v}) \bullet (\vec{u} + \vec{v}) \\
& = & (\vec{u} \bullet \vec{u}) + (\vec{u} \bullet \vec{v}) + (\vec{v} \bullet \vec{u}) + (\vec{v} \bullet \vec{v}) \\
& = & \|\vec{u}\|^2 + 2(\vec{u} \bullet \vec{v}) + \|\vec{v}\|^2 \\
& \leq & \|\vec{u}\|^2 + 2|\vec{u} \bullet \vec{v}| + \|\vec{v}\|^2 \\
& \leq & \|\vec{u}\|^2 + 2\|\vec{u}\|\|\vec{v}\| + \|\vec{v}\|^2 = (\|\vec{u}\| + \|\vec{v}\|)^2
\end{array}$$

Hence,

$$\|\vec{u} + \vec{v}\|^2 \leq (\|\vec{u}\| + \|\vec{v}\|)^2$$

Taking square roots of both sides you obtain 4.9.

It remains to consider when equality occurs. Suppose $\vec{u} = \vec{0}$. Then, $\vec{u} = 0\vec{v}$ and the claim about when equality occurs is verified. The same argument holds if $\vec{v} = \vec{0}$. Therefore, it can be assumed both vectors are nonzero. To get equality in 4.9 above, Theorem 4.29 implies one of the vectors must be a multiple of the other. Say $\vec{v} = k\vec{u}$. If $k < 0$ then equality cannot occur in 4.9 because in this case

$$\vec{u} \bullet \vec{v} = k\|\vec{u}\|^2 < 0 < |k|\,\|\vec{u}\|^2 = |\vec{u} \bullet \vec{v}|$$

Therefore, $k \geq 0$.

To get the other form of the triangle inequality write

$$\vec{u} = \vec{u} - \vec{v} + \vec{v}$$

so

$$\begin{aligned}\|\vec{u}\| &= \|\vec{u} - \vec{v} + \vec{v}\| \\ &\leq \|\vec{u} - \vec{v}\| + \|\vec{v}\|\end{aligned}$$

Therefore,

$$\|\vec{u}\| - \|\vec{v}\| \leq \|\vec{u} - \vec{v}\| \tag{4.11}$$

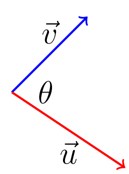
Similarly,

$$\|\vec{v}\| - \|\vec{u}\| \leq \|\vec{v} - \vec{u}\| = \|\vec{u} - \vec{v}\| \tag{4.12}$$

It follows from 4.11 and 4.12 that 4.10 holds. This is because $|\|\vec{u}\| - \|\vec{v}\||$ equals the left side of either 4.11 or 4.12 and either way, $|\|\vec{u}\| - \|\vec{v}\|| \leq \|\vec{u} - \vec{v}\|$. □

4.7.2. The Geometric Significance of the Dot Product

Given two vectors, $\vec{u}$ and $\vec{v}$, the **included angle** is the angle between these two vectors which is given by θ such that $0 \leq \theta \leq \pi$. The dot product can be used to determine the included angle between two vectors. Consider the following picture where θ gives the included angle.



Proposition 4.31: The Dot Product and the Included Angle

Let $\vec{u}$ and $\vec{v}$ be two vectors in $\mathbb{R}^n$, and let θ be the included angle. Then the following equation holds.

$$\vec{u} \bullet \vec{v} = \|\vec{u}\| \|\vec{v}\| \cos \theta$$

In words, the dot product of two vectors equals the product of the magnitude (or length) of the two vectors multiplied by the cosine of the included angle. Note this gives a geometric description of the dot product which does not depend explicitly on the coordinates of the vectors.

Consider the following example.

Example 4.32: Find the Angle Between Two Vectors

Find the angle between the vectors given by

$$\vec{u} = \left[\begin{array}{r} 2 \\ 1 \\ -1 \end{array}\right], \vec{v} = \left[\begin{array}{r} 3 \\ 4 \\ 1 \end{array}\right]$$

Solution. By Proposition 4.31,

$$\vec{u} \bullet \vec{v} = \|\vec{u}\| \|\vec{v}\| \cos\theta$$

Hence,

$$\cos\theta = \frac{\vec{u} \bullet \vec{v}}{\|\vec{u}\| \|\vec{v}\|}$$

First, we can compute $\vec{u} \bullet \vec{v}$. By Definition 4.25, this equals

$$\vec{u} \bullet \vec{v} = (2)(3) + (1)(4) + (-1)(1) = 9$$

Then,

$$\begin{array}{c} \|\vec{u}\| = \sqrt{(2)(2) + (1)(1) + (1)(1)} = \sqrt{6} \\ \|\vec{v}\| = \sqrt{(3)(3) + (4)(4) + (1)(1)} = \sqrt{26} \end{array}$$

Therefore, the cosine of the included angle equals

$$\cos\theta = \frac{9}{\sqrt{26}\sqrt{6}} = 0.7205766...$$

With the cosine known, the angle can be determined by computing the inverse cosine of that angle, giving approximately $\theta = 0.76616$ radians. ◻

Another application of the geometric description of the dot product is in finding the angle between two lines. Typically one would assume that the lines intersect. In some situations, however, it may make sense to ask this question when the lines do not intersect, such as the angle between two object trajectories. In any case we understand it to mean the smallest angle between (any of) their direction vectors. The only subtlety here is that if $\vec{u}$ is a direction vector for a line, then so is any multiple $k\vec{u}$, and thus we will find complementary angles among all angles between direction vectors for two lines, and we simply take the smaller of the two.

Example 4.33: Find the Angle Between Two Lines

Find the angle between the two lines

$$L_1 : \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$$

and

$$L_2 : \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ -3 \end{bmatrix} + s \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

Solution. You can verify that these lines do not intersect, but as discussed above this does not matter and we simply find the smallest angle between any directions vectors for these lines.

To do so we first find the angle between the direction vectors given above:

$$\vec{u} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, \ \vec{v} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

In order to find the angle, we solve the following equation for θ

$$\vec{u} \bullet \vec{v} = \|\vec{u}\|\|\vec{v}\| \cos\theta$$

to obtain $\cos\theta = -\frac{1}{2}$ and since we choose included angles between 0 and π we obtain $\theta = \frac{2\pi}{3}$.

Now the angles between any two direction vectors for these lines will either be $\frac{2\pi}{3}$ or its complement $\phi = \pi - \frac{2\pi}{3} = \frac{\pi}{3}$. We choose the smaller angle, and therefore conclude that the angle between the two lines is $\frac{\pi}{3}$. ♠

We can also use Proposition 4.31 to compute the dot product of two vectors.

Example 4.34: Using Geometric Description to Find a Dot Product

Let $\vec{u}, \vec{v}$ be vectors with $\|\vec{u}\| = 3$ and $\|\vec{v}\| = 4$. Suppose the angle between $\vec{u}$ and $\vec{v}$ is $\pi/3$. Find $\vec{u} \bullet \vec{v}$.

Solution. From the geometric description of the dot product in Proposition 4.31

$$\vec{u} \bullet \vec{v} = (3)(4)\cos(\pi/3) = 3 \times 4 \times 1/2 = 6$$

♠

Two nonzero vectors are said to be **perpendicular**, sometimes also called **orthogonal**, if the included angle is $\pi/2$ radians (90°).

Consider the following proposition.

Proposition 4.35: Perpendicular Vectors

*Let $\vec{u}$ and $\vec{v}$ be nonzero vectors in $\mathbb{R}^n$. Then, $\vec{u}$ and $\vec{v}$ are said to be **perpendicular** exactly when*

$$\vec{u} \bullet \vec{v} = 0$$

Proof. This follows directly from Proposition 4.31. First if the dot product of two nonzero vectors is equal to 0, this tells us that $\cos\theta = 0$ (this is where we need nonzero vectors). Thus $\theta = \pi/2$ and the vectors are perpendicular.

If on the other hand $\vec{v}$ is perpendicular to $\vec{u}$, then the included angle is $\pi/2$ radians. Hence $\cos\theta = 0$ and $\vec{u} \bullet \vec{v} = 0$. □

Consider the following example.

Example 4.36: Determine if Two Vectors are Perpendicular

Determine whether the two vectors,

$$\vec{u} = \left[\begin{array}{r} 2 \\ 1 \\ -1 \end{array}\right], \vec{v} = \left[\begin{array}{r} 1 \\ 3 \\ 5 \end{array}\right]$$

are perpendicular.

Solution. In order to determine if these two vectors are perpendicular, we compute the dot product. This is given by

$$\vec{u} \bullet \vec{v} = (2)(1) + (1)(3) + (-1)(5) = 0$$

Therefore, by Proposition 4.35 these two vectors are perpendicular. □

4.7.3. Projections

In some applications, we wish to write a vector as a sum of two related vectors. Through the concept of projections, we can find these two vectors. First, we explore an important theorem. The result of this theorem will provide our definition of a vector projection.

Theorem 4.37: Vector Projections

Let $\vec{v}$ and $\vec{u}$ be nonzero vectors. Then there exist unique vectors $\vec{v}_{||}$ and $\vec{v}_{\perp}$ such that

$$\vec{v} = \vec{v}_{||} + \vec{v}_{\perp} \tag{4.13}$$

where $\vec{v}_{||}$ is a scalar multiple of $\vec{u}$, and $\vec{v}_{\perp}$ is perpendicular to $\vec{u}$.

Proof. Suppose 4.13 holds and $\vec{v}_{||} = k\vec{u}$. Taking the dot product of both sides of 4.13 with $\vec{u}$ and using $\vec{v}_{\perp} \bullet \vec{u} = 0$, this yields

$$\begin{aligned} \vec{v} \bullet \vec{u} &= (\vec{v}_{||} + \vec{v}_{\perp}) \bullet \vec{u} \\ &= k\vec{u} \bullet \vec{u} + \vec{v}_{\perp} \bullet \vec{u} \\ &= k\|\vec{u}\|^2 \end{aligned}$$

which requires $k = \vec{v} \bullet \vec{u}/\|\vec{u}\|^2$. Thus there can be no more than one vector $\vec{v}_{||}$. It follows $\vec{v}_{\perp}$ must equal $\vec{v} - \vec{v}_{||}$. This verifies there can be no more than one choice for both $\vec{v}_{||}$ and $\vec{v}_{\perp}$ and proves their uniqueness.

Now let

$$\vec{v}_{||} = \frac{\vec{v} \bullet \vec{u}}{\|\vec{u}\|^2}\vec{u}$$

and let

$$\vec{v}_{\perp} = \vec{v} - \vec{v}_{||} = \vec{v} - \frac{\vec{v} \bullet \vec{u}}{\|\vec{u}\|^2}\vec{u}$$

Then $\vec{v}_{||} = k\vec{u}$ where $k = \frac{\vec{v} \bullet \vec{u}}{\|\vec{u}\|^2}$. It only remains to verify $\vec{v}_{\perp} \bullet \vec{u} = 0$. But

$$\begin{aligned} \vec{v}_{\perp} \bullet \vec{u} &= \vec{v} \bullet \vec{u} - \frac{\vec{v} \bullet \vec{u}}{\|\vec{u}\|^2}\vec{u} \bullet \vec{u} \\ &= \vec{v} \bullet \vec{u} - \vec{v} \bullet \vec{u} \\ &= 0 \end{aligned}$$

□

The vector $\vec{v}_{||}$ in Theorem 4.37 is called the **projection** of $\vec{v}$ onto $\vec{u}$ and is denoted by

$$\vec{v}_{||} = \mathrm{proj}_{\vec{u}}(\vec{v})$$

We now make a formal definition of the vector projection.

Definition 4.38: Vector Projection

Let $\vec{u}$ and $\vec{v}$ be vectors. Then, the ***projection of*** $\vec{v}$ ***onto*** $\vec{u}$ *is given by*

$$\mathrm{proj}_{\vec{u}}(\vec{v}) = \left(\frac{\vec{v} \bullet \vec{u}}{\vec{u} \bullet \vec{u}}\right)\vec{u} = \frac{\vec{v} \bullet \vec{u}}{\|\vec{u}\|^2}\vec{u}$$

Consider the following example of a projection.

Example 4.39: Find the Projection of One Vector Onto Another

Find $\mathrm{proj}_{\vec{u}}(\vec{v})$ *if*

$$\vec{u} = \begin{bmatrix} 2 \\ 3 \\ -4 \end{bmatrix}, \vec{v} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Solution. We can use the formula provided in Definition 4.38 to find $\text{proj}_{\vec{u}}(\vec{v})$. First, compute $\vec{v} \bullet \vec{u}$. This is given by

$$\begin{aligned}
\left[\begin{array}{r} 1 \\ -2 \\ 1 \end{array}\right] \bullet \left[\begin{array}{r} 2 \\ 3 \\ -4 \end{array}\right] &= (2)(1) + (3)(-2) + (-4)(1) \\
&= 2 - 6 - 4 \\
&= -8
\end{aligned}$$

Similarly, $\vec{u} \bullet \vec{u}$ is given by

$$\begin{aligned}
\left[\begin{array}{r} 2 \\ 3 \\ -4 \end{array}\right] \bullet \left[\begin{array}{r} 2 \\ 3 \\ -4 \end{array}\right] &= (2)(2) + (3)(3) + (-4)(-4) \\
&= 4 + 9 + 16 \\
&= 29
\end{aligned}$$

Therefore, the projection is equal to

$$\begin{aligned}
\text{proj}_{\vec{u}}(\vec{v}) &= -\frac{8}{29} \left[\begin{array}{r} 2 \\ 3 \\ -4 \end{array}\right] \\
&= \left[\begin{array}{r} -\frac{16}{29} \\ -\frac{24}{29} \\ \frac{32}{29} \end{array}\right]
\end{aligned}$$

□

We will conclude this section with an important application of projections. Suppose a line L and a point P are given such that P is not contained in L. Through the use of projections, we can determine the shortest distance from P to L.

Example 4.40: Shortest Distance from a Point to a Line

Let $P = (1,3,5)$ be a point in $\mathbb{R}^3$, and let L be the line which goes through point $P_0 = (0,4,-2)$ with direction vector $\vec{d} = \left[\begin{array}{r} 2 \\ 1 \\ 2 \end{array}\right]$. Find the shortest distance from P to the line L, and find the point Q on L that is closest to P.

Solution. In order to determine the shortest distance from P to L, we will first find the vector $\overrightarrow{P_0P}$ and then find the projection of this vector onto L. The vector $\overrightarrow{P_0P}$ is given by

$$\left[\begin{array}{r} 1 \\ 3 \\ 5 \end{array}\right] - \left[\begin{array}{r} 0 \\ 4 \\ -2 \end{array}\right] = \left[\begin{array}{r} 1 \\ -1 \\ 7 \end{array}\right]$$

Then, if Q is the point on L closest to P, it follows that

$$\begin{aligned}
\overrightarrow{P_0Q} &= \text{proj}_{\vec{d}}\overrightarrow{P_0P} \\
&= \left(\frac{\overrightarrow{P_0P} \bullet \vec{d}}{\|\vec{d}\|^2} \right) \vec{d} \\
&= \frac{15}{9} \left[\begin{array}{c} 2 \\ 1 \\ 2 \end{array} \right] \\
&= \frac{5}{3} \left[\begin{array}{c} 2 \\ 1 \\ 2 \end{array} \right]
\end{aligned}$$

Now, the distance from P to L is given by

$$\|\overrightarrow{QP}\| = \|\overrightarrow{P_0P} - \overrightarrow{P_0Q}\| = \sqrt{26}$$

The point Q is found by adding the vector $\overrightarrow{P_0Q}$ to the position vector $\overrightarrow{0P_0}$ for P_0 as follows

$$\left[\begin{array}{r} 0 \\ 4 \\ -2 \end{array} \right] + \frac{5}{3} \left[\begin{array}{c} 2 \\ 1 \\ 2 \end{array} \right] = \left[\begin{array}{r} \frac{10}{3} \\ \frac{17}{3} \\ \frac{4}{3} \end{array} \right]$$

Therefore, $Q = (\frac{10}{3}, \frac{17}{3}, \frac{4}{3})$.

♢

Exercises

Exercise 4.7.1 *Find* $\left[\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \right] \bullet \left[\begin{array}{c} 2 \\ 0 \\ 1 \\ 3 \end{array} \right]$.

Exercise 4.7.2 *Use the formula given in Proposition 4.31 to verify the Cauchy Schwarz inequality and to show that equality occurs if and only if one of the vectors is a scalar multiple of the other.*

Exercise 4.7.3 *For $\vec{u}, \vec{v}$ vectors in $\mathbb{R}^3$, define the product, $\vec{u} * \vec{v} = u_1v_1 + 2u_2v_2 + 3u_3v_3$. Show the axioms for a dot product all hold for this product. Prove*

$$\|\vec{u} * \vec{v}\| \leq (\vec{u} * \vec{u})^{1/2} (\vec{v} * \vec{v})^{1/2}$$

Exercise 4.7.4 *Let $\vec{a}, \vec{b}$ be vectors. Show that* $\left(\vec{a} \bullet \vec{b}\right) = \frac{1}{4}\left(\|\vec{a}+\vec{b}\|^2 - \|\vec{a}-\vec{b}\|^2\right)$.

Exercise 4.7.5 *Using the axioms of the dot product, prove the parallelogram identity:*

$$\|\vec{a}+\vec{b}\|^2 + \|\vec{a}-\vec{b}\|^2 = 2\|\vec{a}\|^2 + 2\|\vec{b}\|^2$$

Exercise 4.7.6 *Let A be a real $m \times n$ matrix and let $\vec{u} \in \mathbb{R}^n$ and $\vec{v} \in \mathbb{R}^m$. Show $A\vec{u} \bullet \vec{v} = \vec{u} \bullet A^T\vec{v}$.* ***Hint:*** *Use the definition of matrix multiplication to do this.*

Exercise 4.7.7 *Use the result of Problem 4.7.6 to verify directly that $(AB)^T = B^T A^T$ without making any reference to subscripts.*

Exercise 4.7.8 *Find the angle between the vectors*

$$\vec{u} = \left[\begin{array}{r} 3 \\ -1 \\ -1 \end{array}\right], \vec{v} = \left[\begin{array}{r} 1 \\ 4 \\ 2 \end{array}\right]$$

Exercise 4.7.9 *Find the angle between the vectors*

$$\vec{u} = \left[\begin{array}{r} 1 \\ -2 \\ 1 \end{array}\right], \vec{v} = \left[\begin{array}{r} 1 \\ 2 \\ -7 \end{array}\right]$$

Exercise 4.7.10 *Find $\mathrm{proj}_{\vec{v}}(\vec{w})$ where* $\vec{w} = \left[\begin{array}{r} 1 \\ 0 \\ -2 \end{array}\right]$ *and* $\vec{v} = \left[\begin{array}{r} 1 \\ 2 \\ 3 \end{array}\right]$.

Exercise 4.7.11 *Find $\mathrm{proj}_{\vec{v}}(\vec{w})$ where* $\vec{w} = \left[\begin{array}{r} 1 \\ 2 \\ -2 \end{array}\right]$ *and* $\vec{v} = \left[\begin{array}{r} 1 \\ 0 \\ 3 \end{array}\right]$.

Exercise 4.7.12 *Find $\mathrm{proj}_{\vec{v}}(\vec{w})$ where* $\vec{w} = \left[\begin{array}{r} 1 \\ 2 \\ -2 \\ 1 \end{array}\right]$ *and* $\vec{v} = \left[\begin{array}{r} 1 \\ 2 \\ 3 \\ 0 \end{array}\right]$.

Exercise 4.7.13 *Let $P = (1,2,3)$ be a point in $\mathbb{R}^3$. Let L be the line through the point $P_0 = (1,4,5)$ with direction vector* $\vec{d} = \left[\begin{array}{r} 1 \\ -1 \\ 1 \end{array}\right]$. *Find the shortest distance from P to L, and find the point Q on L that is closest to P.*

Exercise 4.7.14 *Let $P = (0,2,1)$ be a point in $\mathbb{R}^3$. Let L be the line through the point $P_0 = (1,1,1)$ with direction vector* $\vec{d} = \left[\begin{array}{r} 3 \\ 0 \\ 1 \end{array}\right]$. *Find the shortest distance from P to L, and find the point Q on L that is closest to P.*

Exercise 4.7.15 *Does it make sense to speak of* $proj_{\vec{0}}(\vec{w})$?

Exercise 4.7.16 *Prove the Cauchy Schwarz inequality in* $\mathbb{R}^n$ *as follows. For* $\vec{u}, \vec{v}$ *vectors, consider*

$$(\vec{w} - proj_{\vec{v}}\vec{w}) \bullet (\vec{w} - proj_{\vec{v}}\vec{w}) \geq 0$$

Simplify using the axioms of the dot product and then put in the formula for the projection. Notice that this expression equals 0 *and you get equality in the Cauchy Schwarz inequality if and only if* $\vec{w} = proj_{\vec{v}}\vec{w}$. *What is the geometric meaning of* $\vec{w} = proj_{\vec{v}}\vec{w}$?

Exercise 4.7.17 *Let* $\vec{v}, \vec{w}$ $\vec{u}$ *be vectors. Show that* $(\vec{w} + \vec{u})_{\perp} = \vec{w}_{\perp} + \vec{u}_{\perp}$ *where* $\vec{w}_{\perp} = \vec{w} - proj_{\vec{v}}(\vec{w})$.

Exercise 4.7.18 *Show that*

$$(\vec{v} - proj_{\vec{u}}(\vec{v}), \vec{u}) = (\vec{v} - proj_{\vec{u}}(\vec{v})) \bullet \vec{u} = 0$$

and conclude every vector in $\mathbb{R}^n$ *can be written as the sum of two vectors, one which is perpendicular and one which is parallel to the given vector.*

4.8 Planes in $\mathbb{R}^n$

Outcomes

A. *Find the vector and scalar equations of a plane.*

Much like the above discussion with lines, vectors can be used to determine planes in $\mathbb{R}^n$. Given a vector $\vec{n}$ in $\mathbb{R}^n$ and a point P_0, it is possible to find a **unique** plane which contains P_0 and is perpendicular to the given vector.

Definition 4.41: Normal Vector

Let $\vec{n}$ be a nonzero vector in $\mathbb{R}^n$. Then $\vec{n}$ is called a ***normal vector*** *to a plane if and only if*

$$\vec{n} \bullet \vec{v} = 0$$

for every vector $\vec{v}$ in the plane.

In other words, we say that $\vec{n}$ is orthogonal (perpendicular) to every vector in the plane.

Consider now a plane with normal vector given by $\vec{n}$, and containing a point P_0. Notice that this plane is unique. If P is an arbitrary point on this plane, then by definition the normal vector is orthogonal to the vector between P_0 and P. Letting $\overrightarrow{0P}$ and $\overrightarrow{0P_0}$ be the position vectors of points P and P_0 respectively, it follows that

$$\vec{n} \bullet (\overrightarrow{0P} - \overrightarrow{0P_0}) = 0$$

or

$$\vec{n} \bullet \overrightarrow{P_0P} = 0$$

The first of these equations gives the **vector equation** of the plane.

Definition 4.42: Vector Equation of a Plane

Let $\vec{n}$ be the normal vector for a plane which contains a point P_0. If P is an arbitrary point on this plane, then the ***vector equation*** *of the plane is given by*

$$\vec{n} \bullet (\overrightarrow{0P} - \overrightarrow{0P_0}) = 0$$

Notice that this equation can be used to determine if a point P is contained in a certain plane.

Example 4.43: A Point in a Plane

Let $\vec{n} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ *be the normal vector for a plane which contains the point* $P_0 = (2, 1, 4)$.
Determine if the point $P = (5, 4, 1)$ *is contained in this plane.*

Solution. By Definition 4.42, P is a point in the plane if it satisfies the equation

$$\vec{n} \bullet (\overrightarrow{0P} - \overrightarrow{0P_0}) = 0$$

Given the above $\vec{n}$, P_0, and P, this equation becomes

$$\begin{aligned} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \bullet \left(\begin{bmatrix} 5 \\ 4 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} \right) &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \bullet \left(\begin{bmatrix} 3 \\ 3 \\ -3 \end{bmatrix} \right) \\ &= 3 + 6 - 9 = 0 \end{aligned}$$

Therefore $P = (5, 4, 1)$ is contained in the plane.

□

Suppose $\vec{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, $P = (x, y, z)$ and $P_0 = (x_0, y_0, z_0)$.
Then

$$\begin{aligned} \vec{n} \bullet (\overrightarrow{0P} - \overrightarrow{0P_0}) &= 0 \\ \begin{bmatrix} a \\ b \\ c \end{bmatrix} \bullet \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \right) &= 0 \\ \begin{bmatrix} a \\ b \\ c \end{bmatrix} \bullet \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix} &= 0 \\ a(x - x_0) + b(y - y_0) + c(z - z_0) &= 0 \end{aligned}$$

We can also write this equation as

$$ax + by + cz = ax_0 + by_0 + cz_0$$

Notice that since P_0 is given, $ax_0 + by_0 + cz_0$ is a known scalar, which we can call d. This equation becomes

$$ax + by + cz = d$$

Definition 4.44: Scalar Equation of a Plane

Let $\vec{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ *be the normal vector for a plane which contains the point* $P_0 = (x_0, y_0, z_0)$*.Then if* $P = (x,y,z)$ *is an arbitrary point on the plane, the* ***scalar equation*** *of the plane is given by*

$$ax + by + cz = d$$

where $a,b,c,d \in \mathbb{R}$ *and* $d = ax_0 + by_0 + cz_0$*.*

Consider the following equation.

Example 4.45: Finding the Equation of a Plane

Find an equation of the plane containing $P_0 = (3,-2,5)$ *and orthogonal to* $\vec{n} = \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix}$.

Solution. The above vector $\vec{n}$ is the normal vector for this plane. Using Definition 4.42, we can determine the vector equation for this plane.

$$\begin{aligned} \vec{n} \bullet (\overrightarrow{0P} - \overrightarrow{0P_0}) &= 0 \\ \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix} \bullet \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} 3 \\ -2 \\ 5 \end{bmatrix} \right) &= 0 \\ \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix} \bullet \begin{bmatrix} x-3 \\ y+2 \\ z-5 \end{bmatrix} &= 0 \end{aligned}$$

Using Definition 4.44, we can determine the scalar equation of the plane.

$$-2x + 4y + 1z = -2(3) + 4(-2) + 1(5) = -9$$

Hence, the vector equation of the plane is

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

and the scalar equation is

$$-2x + 4y + 1z = -9$$

$\square$

Suppose a point P is not contained in a given plane. We are then interested in the shortest distance from that point P to the given plane. Consider the following example.

Example 4.46: Shortest Distance From a Point to a Plane

Find the shortest distance from the point $P = (3,2,3)$ to the plane given by $2x + y + 2z = 2$, and find the point Q on the plane that is closest to P.

Solution. Pick an arbitrary point P_0 on the plane. Then, it follows that

$$\overrightarrow{QP} = \mathrm{proj}_{\vec{n}} \overrightarrow{P_0P}$$

and $\|\overrightarrow{QP}\|$ is the shortest distance from P to the plane. Further, the vector $\overrightarrow{0Q} = \overrightarrow{0P} - \overrightarrow{QP}$ gives the necessary point Q.

From the above scalar equation, we have that $\vec{n} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$. Now, choose $P_0 = (1,0,0)$ so that $\vec{n} \bullet \overrightarrow{0P} = 2 = d$. Then, $\overrightarrow{P_0P} = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$.

Next, compute $\overrightarrow{QP} = \mathrm{proj}_{\vec{n}} \overrightarrow{P_0P}$.

$$\begin{aligned} \overrightarrow{QP} &= \mathrm{proj}_{\vec{n}} \overrightarrow{P_0P} \\ &= \left(\frac{\overrightarrow{P_0P} \bullet \vec{n}}{\|\vec{n}\|^2} \right) \vec{n} \\ &= \frac{12}{9} \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \\ &= \frac{4}{3} \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \end{aligned}$$

Then, $\|\overrightarrow{QP}\| = 4$ so the shortest distance from P to the plane is 4.

Next, to find the point Q on the plane which is closest to P we have

$$\begin{aligned} \overrightarrow{0Q} &= \overrightarrow{0P} - \overrightarrow{QP} \\ &= \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} - \frac{4}{3} \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \end{aligned}$$

Therefore, $Q = (\frac{1}{3}, \frac{2}{3}, \frac{1}{3})$. ♢

4.9 The Cross Product

Outcomes

A. *Compute the cross product and box product of vectors in $\mathbb{R}^3$.*

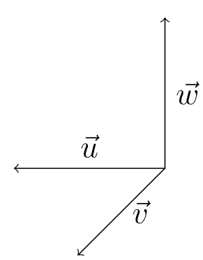
Recall that the dot product is one of two important products for vectors. The second type of product for vectors is called the cross product. It is important to note that the cross product is only defined in $\mathbb{R}^3$. First we discuss the geometric meaning and then a description in terms of coordinates is given, both of which are important. The geometric description is essential in order to understand the applications to physics and geometry while the coordinate description is necessary to compute the cross product.

Consider the following definition.

Definition 4.47: Right Hand System of Vectors

Three vectors, $\vec{u}, \vec{v}, \vec{w}$ form a right hand system if when you extend the fingers of your right hand along the direction of vector $\vec{u}$ and close them in the direction of $\vec{v}$, the thumb points roughly in the direction of $\vec{w}$.

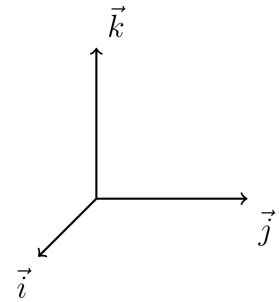
For an example of a right handed system of vectors, see the following picture.



In this picture the vector $\vec{w}$ points upwards from the plane determined by the other two vectors. Point the fingers of your right hand along $\vec{u}$, and close them in the direction of $\vec{v}$. Notice that if you extend the thumb on your right hand, it points in the direction of $\vec{w}$.

You should consider how a right hand system would differ from a left hand system. Try using your left hand and you will see that the vector $\vec{w}$ would need to point in the opposite direction.

Notice that the special vectors, $\vec{i}, \vec{j}, \vec{k}$ will always form a right handed system. If you extend the fingers of your right hand along $\vec{i}$ and close them in the direction $\vec{j}$, the thumb points in the direction of $\vec{k}$.



The following is the geometric description of the cross product. Recall that the dot product of two vectors results in a scalar. In contrast, the cross product results in a vector, as the product gives a direction as well as magnitude.

Definition 4.48: Geometric Definition of Cross Product

*Let $\vec{u}$ and $\vec{v}$ be two vectors in $\mathbb{R}^3$. Then the **cross product**, written $\vec{u} \times \vec{v}$, is defined by the following two rules.*

1. *Its length is $\|\vec{u} \times \vec{v}\| = \|\vec{u}\|\|\vec{v}\| \sin\theta$,*
   where θ is the included angle between $\vec{u}$ and $\vec{v}$.

2. *It is perpendicular to both $\vec{u}$ and $\vec{v}$, that is $(\vec{u} \times \vec{v}) \bullet \vec{u} = 0$, $(\vec{u} \times \vec{v}) \bullet \vec{v} = 0$, and $\vec{u}, \vec{v}, \vec{u} \times \vec{v}$ form a right hand system.*

The cross product of the special vectors $\vec{i}, \vec{j}, \vec{k}$ is as follows.

$$\begin{array}{ll} \vec{i} \times \vec{j} = \vec{k} & \vec{j} \times \vec{i} = -\vec{k} \\ \vec{k} \times \vec{i} = \vec{j} & \vec{i} \times \vec{k} = -\vec{j} \\ \vec{j} \times \vec{k} = \vec{i} & \vec{k} \times \vec{j} = -\vec{i} \end{array}$$

With this information, the following gives the coordinate description of the cross product.

Recall that the vector $\vec{u} = \left[\begin{array}{ccc} u_1 & u_2 & u_3 \end{array} \right]^T$ can be written in terms of $\vec{i}, \vec{j}, \vec{k}$ as $\vec{u} = u_1\vec{i} + u_2\vec{j} + u_3\vec{k}$.

Proposition 4.49: Coordinate Description of Cross Product

Let $\vec{u} = u_1\vec{i} + u_2\vec{j} + u_3\vec{k}$ and $\vec{v} = v_1\vec{i} + v_2\vec{j} + v_3\vec{k}$ be two vectors. Then

$$\begin{array}{c} \vec{u} \times \vec{v} = (u_2v_3 - u_3v_2)\,\vec{i} - (u_1v_3 - u_3v_1)\,\vec{j} + \\ + (u_1v_2 - u_2v_1)\,\vec{k} \end{array} \tag{4.14}$$

Writing $\vec{u} \times \vec{v}$ in the usual way, it is given by

$$\vec{u} \times \vec{v} = \left[\begin{array}{r} u_2v_3 - u_3v_2 \\ -(u_1v_3 - u_3v_1) \\ u_1v_2 - u_2v_1 \end{array} \right]$$

We now prove this proposition.

Proof. From the above table and the properties of the cross product listed,

$$\begin{aligned}
\vec{u} \times \vec{v} &= \left(u_1\vec{i} + u_2\vec{j} + u_3\vec{k}\right) \times \left(v_1\vec{i} + v_2\vec{j} + v_3\vec{k}\right) \\
&= u_1v_2\vec{i} \times \vec{j} + u_1v_3\vec{i} \times \vec{k} + u_2v_1\vec{j} \times \vec{i} + u_2v_3\vec{j} \times \vec{k} + +u_3v_1\vec{k} \times \vec{i} + u_3v_2\vec{k} \times \vec{j} \\
&= u_1v_2\vec{k} - u_1v_3\vec{j} - u_2v_1\vec{k} + u_2v_3\vec{i} + u_3v_1\vec{j} - u_3v_2\vec{i} \\
&= (u_2v_3 - u_3v_2)\,\vec{i} + (u_3v_1 - u_1v_3)\,\vec{j} + (u_1v_2 - u_2v_1)\,\vec{k}
\end{aligned} \tag{4.15}$$

□

There is another version of 4.14 which may be easier to remember. We can express the cross product as the determinant of a matrix, as follows.

$$\vec{u} \times \vec{v} = \left| \begin{array}{ccc} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{array} \right| \tag{4.16}$$

Expanding the determinant along the top row yields

$$\vec{i}\,(-1)^{1+1} \left| \begin{array}{cc} u_2 & u_3 \\ v_2 & v_3 \end{array} \right| + \vec{j}\,(-1)^{2+1} \left| \begin{array}{cc} u_1 & u_3 \\ v_1 & v_3 \end{array} \right| + \vec{k}\,(-1)^{3+1} \left| \begin{array}{cc} u_1 & u_2 \\ v_1 & v_2 \end{array} \right|$$

$$= \vec{i} \left| \begin{array}{cc} u_2 & u_3 \\ v_2 & v_3 \end{array} \right| - \vec{j} \left| \begin{array}{cc} u_1 & u_3 \\ v_1 & v_3 \end{array} \right| + \vec{k} \left| \begin{array}{cc} u_1 & u_2 \\ v_1 & v_2 \end{array} \right|$$

Expanding these determinants leads to

$$(u_2v_3 - u_3v_2)\,\vec{i} - (u_1v_3 - u_3v_1)\,\vec{j} + (u_1v_2 - u_2v_1)\,\vec{k}$$

which is the same as 4.15.

The cross product satisfies the following properties.

Proposition 4.50: Properties of the Cross Product

Let $\vec{u}, \vec{v}, \vec{w}$ be vectors in $\mathbb{R}^3$, and k a scalar. Then, the following properties of the cross product hold.

1. $\vec{u} \times \vec{v} = -\left(\vec{v} \times \vec{u}\right)$, and $\vec{u} \times \vec{u} = \vec{0}$
2. $(k\vec{u}) \times \vec{v} = k\left(\vec{u} \times \vec{v}\right) = \vec{u} \times (k\vec{v})$
3. $\vec{u} \times (\vec{v} + \vec{w}) = \vec{u} \times \vec{v} + \vec{u} \times \vec{w}$
4. $(\vec{v} + \vec{w}) \times \vec{u} = \vec{v} \times \vec{u} + \vec{w} \times \vec{u}$

Proof. Formula 1. follows immediately from the definition. The vectors $\vec{u} \times \vec{v}$ and $\vec{v} \times \vec{u}$ have the same magnitude, $|\vec{u}|\,|\vec{v}|\sin\theta$, and an application of the right hand rule shows they have opposite direction.

Formula 2. is proven as follows. If k is a non-negative scalar, the direction of $(k\vec{u}) \times \vec{v}$ is the same as the direction of $\vec{u} \times \vec{v}, k(\vec{u} \times \vec{v})$ and $\vec{u} \times (k\vec{v})$. The magnitude is k times the magnitude of $\vec{u} \times \vec{v}$ which is the same as the magnitude of $k(\vec{u} \times \vec{v})$ and $\vec{u} \times (k\vec{v})$. Using this yields equality in 2. In the case where $k < 0$, everything works the same way except the vectors are all pointing in the opposite direction and you must multiply by $|k|$ when comparing their magnitudes.

The distributive laws, 3. and 4., are much harder to establish. For now, it suffices to notice that if we know that 3. is true, 4. follows. Thus, assuming 3., and using 1.,

$$\begin{aligned}(\vec{v} + \vec{w}) \times \vec{u} &= -\vec{u} \times (\vec{v} + \vec{w}) \\ &= -(\vec{u} \times \vec{v} + \vec{u} \times \vec{w}) \\ &= \vec{v} \times \vec{u} + \vec{w} \times \vec{u}\end{aligned}$$

□

We will now look at an example of how to compute a cross product.

Example 4.51: Find a Cross Product

Find $\vec{u} \times \vec{v}$ for the following vectors

$$\vec{u} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \vec{v} = \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix}$$

Solution. Note that we can write $\vec{u}, \vec{v}$ in terms of the special vectors $\vec{i}, \vec{j}, \vec{k}$ as

$$\begin{aligned}\vec{u} &= \vec{i} - \vec{j} + 2\vec{k} \\ \vec{v} &= 3\vec{i} - 2\vec{j} + \vec{k}\end{aligned}$$

We will use the equation given by 4.16 to compute the cross product.

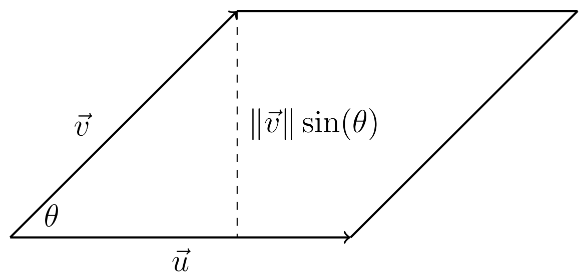
$$\vec{u} \times \vec{v} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 1 & -1 & 2 \\ 3 & -2 & 1 \end{vmatrix} = \begin{vmatrix} -1 & 2 \\ -2 & 1 \end{vmatrix} \vec{i} - \begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} \vec{j} + \begin{vmatrix} 1 & -1 \\ 3 & -2 \end{vmatrix} \vec{k} = 3\vec{i} + 5\vec{j} + \vec{k}$$

We can write this result in the usual way, as

$$\vec{u} \times \vec{v} = \begin{bmatrix} 3 \\ 5 \\ 1 \end{bmatrix}$$

□

An important geometrical application of the cross product is as follows. The size of the cross product, $\|\vec{u} \times \vec{v}\|$, is the area of the parallelogram determined by $\vec{u}$ and $\vec{v}$, as shown in the following picture.



We examine this concept in the following example.

Example 4.52: Area of a Parallelogram

Find the area of the parallelogram determined by the vectors $\vec{u}$ *and* $\vec{v}$ *given by*

$$\vec{u} = \left[\begin{array}{r} 1 \\ -1 \\ 2 \end{array}\right], \vec{v} = \left[\begin{array}{r} 3 \\ -2 \\ 1 \end{array}\right]$$

Solution. Notice that these vectors are the same as the ones given in Example 4.51. Recall from the geometric description of the cross product, that the area of the parallelogram is simply the magnitude of $\vec{u} \times \vec{v}$. From Example 4.51, $\vec{u} \times \vec{v} = 3\vec{i} + 5\vec{j} + \vec{k}$. We can also write this as

$$\vec{u} \times \vec{v} = \left[\begin{array}{r} 3 \\ 5 \\ 1 \end{array}\right]$$

Thus the area of the parallelogram is

$$\|\vec{u} \times \vec{v}\| = \sqrt{(3)(3) + (5)(5) + (1)(1)} = \sqrt{9 + 25 + 1} = \sqrt{35}$$

$\square$

We can also use this concept to find the area of a triangle. Consider the following example.

Example 4.53: Area of Triangle

Find the area of the triangle determined by the points $(1,2,3), (0,2,5), (5,1,2)$

Solution. This triangle is obtained by connecting the three points with lines. Picking $(1,2,3)$ as a starting point, there are two displacement vectors, $\left[\begin{array}{rrr} -1 & 0 & 2 \end{array}\right]^T$ and $\left[\begin{array}{rrr} 4 & -1 & -1 \end{array}\right]^T$. Notice that if we add either of these vectors to the position vector of the starting point, the result is the position vectors of the other two points. Now, the area of the triangle is half the area of the parallelogram determined by $\left[\begin{array}{rrr} -1 & 0 & 2 \end{array}\right]^T$ and $\left[\begin{array}{rrr} 4 & -1 & -1 \end{array}\right]^T$. The required cross product is given by

$$\left[\begin{array}{r} -1 \\ 0 \\ 2 \end{array}\right] \times \left[\begin{array}{r} 4 \\ -1 \\ -1 \end{array}\right] = \left[\begin{array}{rrr} 2 & 7 & 1 \end{array}\right]$$

Taking the size of this vector gives the area of the parallelogram, given by

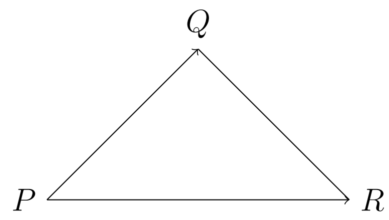
$$\sqrt{(2)(2)+(7)(7)+(1)(1)} = \sqrt{4+49+1} = \sqrt{54}$$

Hence the area of the triangle is $\frac{1}{2}\sqrt{54} = \frac{3}{2}\sqrt{6}$. ◻

In general, if you have three points in $\mathbb{R}^3$, P, Q, R, the area of the triangle is given by

$$\frac{1}{2}\|\overrightarrow{PQ} \times \overrightarrow{PR}\|$$

Recall that $\overrightarrow{PQ}$ is the vector running from point P to point Q.



In the next section, we explore another application of the cross product.

4.9.1. The Box Product

Recall that we can use the cross product to find the the area of a parallelogram. It follows that we can use the cross product together with the dot product to find the volume of a parallelepiped.

We begin with a definition.

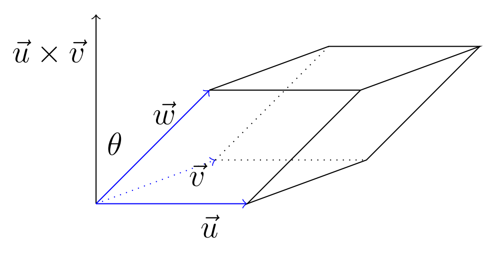
Definition 4.54: Parallelepiped

A parallelepiped determined by the three vectors, $\vec{u}, \vec{v}$, and $\vec{w}$ consists of

$$\{r\vec{u} + s\vec{v} + t\vec{w} : r, s, t \in [0,1]\}$$

That is, if you pick three numbers, r, s, and t each in $[0,1]$ and form $r\vec{u} + s\vec{v} + t\vec{w}$ then the collection of all such points makes up the parallelepiped determined by these three vectors.

The following is an example of a parallelepiped.



Notice that the base of the parallelepiped is the parallelogram determined by the vectors $\vec{u}$ and $\vec{v}$. Therefore, its area is equal to $\|\vec{u} \times \vec{v}\|$. The height of the parallelepiped is $\|\vec{w}\| \cos\theta$ where θ is the angle shown in the picture between $\vec{w}$ and $\vec{u} \times \vec{v}$. The volume of this parallelepiped is the area of the base times the height which is just

$$\|\vec{u} \times \vec{v}\| \|\vec{w}\| \cos\theta = \vec{u} \times \vec{v} \bullet \vec{w}$$

This expression is known as the box product and is sometimes written as $[\vec{u}, \vec{v}, \vec{w}]$. You should consider what happens if you interchange the $\vec{v}$ with the $\vec{w}$ or the $\vec{u}$ with the $\vec{w}$. You can see geometrically from drawing pictures that this merely introduces a minus sign. In any case the box product of three vectors always equals either the volume of the parallelepiped determined by the three vectors or else -1 times this volume.

Proposition 4.55: The Box Product

Let $\vec{u}, \vec{v}, \vec{w}$ be three vectors in $\mathbb{R}^n$ that define a parallelepiped. Then the volume of the parallelepiped is the absolute value of the box product, given by

$$|\vec{u} \times \vec{v} \bullet \vec{w}|$$

Consider an example of this concept.

Example 4.56: Volume of a Parallelepiped

Find the volume of the parallelepiped determined by the vectors

$$\vec{u} = \left[\begin{array}{r} 1 \\ 2 \\ -5 \end{array}\right], \vec{v} = \left[\begin{array}{r} 1 \\ 3 \\ -6 \end{array}\right], \vec{w} = \left[\begin{array}{r} 3 \\ 2 \\ 3 \end{array}\right]$$

Solution. According to the above discussion, pick any two of these vectors, take the cross product and then take the dot product of this with the third of these vectors. The result will be either the desired volume or -1 times the desired volume. Therefore by taking the absolute value of the result, we obtain the volume.

We will take the cross product of $\vec{u}$ and $\vec{v}$. This is given by

$$\vec{u} \times \vec{v} = \left[\begin{array}{r} 1 \\ 2 \\ -5 \end{array}\right] \times \left[\begin{array}{r} 1 \\ 3 \\ -6 \end{array}\right]$$

$$= \left|\begin{array}{rrr} \vec{i} & \vec{j} & \vec{k} \\ 1 & 2 & -5 \\ 1 & 3 & -6 \end{array}\right| = 3\vec{i} + \vec{j} + \vec{k} = \left[\begin{array}{r} 3 \\ 1 \\ 1 \end{array}\right]$$

Now take the dot product of this vector with $\vec{w}$ which yields

$$\begin{aligned}
(\vec{u} \times \vec{v}) \bullet \vec{w} &= \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} \bullet \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} \\
&= \left(3\vec{i} + \vec{j} + \vec{k}\right) \bullet \left(3\vec{i} + 2\vec{j} + 3\vec{k}\right) \\
&= 9 + 2 + 3 \\
&= 14
\end{aligned}$$

This shows the volume of this parallelepiped is 14 cubic units. ∎

There is a fundamental observation which comes directly from the geometric definitions of the cross product and the dot product.

Proposition 4.57: Order of the Product

Let $\vec{u}, \vec{v}$, and $\vec{w}$ be vectors. Then $(\vec{u} \times \vec{v}) \bullet \vec{w} = \vec{u} \bullet (\vec{v} \times \vec{w})$.

Proof. This follows from observing that either $(\vec{u} \times \vec{v}) \bullet \vec{w}$ and $\vec{u} \bullet (\vec{v} \times \vec{w})$ both give the volume of the parallelepiped or they both give -1 times the volume. ∎

Recall that we can express the cross product as the determinant of a particular matrix. It turns out that the same can be done for the box product. Suppose you have three vectors, $\vec{u} = \begin{bmatrix} a & b & c \end{bmatrix}^T, \vec{v} = \begin{bmatrix} d & e & f \end{bmatrix}^T$, and $\vec{w} = \begin{bmatrix} g & h & i \end{bmatrix}^T$. Then the box product $\vec{u} \bullet \vec{v} \times \vec{w}$ is given by the following.

$$\begin{aligned}
\vec{u} \bullet \vec{v} \times \vec{w} &= \begin{bmatrix} a \\ b \\ c \end{bmatrix} \bullet \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ d & e & f \\ g & h & i \end{vmatrix} \\
&= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\
&= \det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}
\end{aligned}$$

To take the box product, you can simply take the determinant of the matrix which results by letting the rows be the components of the given vectors in the order in which they occur in the box product.

This follows directly from the definition of the cross product given above and the way we expand determinants. Thus the volume of a parallelepiped determined by the vectors $\vec{u}, \vec{v}, \vec{w}$ is just the absolute value of the above determinant.

Exercises

Exercise 4.9.1 *Show that if* $\vec{a} \times \vec{u} = \vec{0}$ *for any unit vector* $\vec{u}$, *then* $\vec{a} = \vec{0}$.

Exercise 4.9.2 *Find the area of the triangle determined by the three points,* $(1,2,3), (4,2,0)$ *and* $(-3,2,1)$.

Exercise 4.9.3 *Find the area of the triangle determined by the three points,* $(1,0,3), (4,1,0)$ *and* $(-3,1,1)$.

Exercise 4.9.4 *Find the area of the triangle determined by the three points,* $(1,2,3), (2,3,4)$ *and* $(3,4,5)$. *Did something interesting happen here? What does it mean geometrically?*

Exercise 4.9.5 *Find the area of the parallelogram determined by the vectors* $\left[\begin{array}{c} 1 \\ 2 \\ 3 \end{array}\right], \left[\begin{array}{r} 3 \\ -2 \\ 1 \end{array}\right]$.

Exercise 4.9.6 *Find the area of the parallelogram determined by the vectors* $\left[\begin{array}{c} 1 \\ 0 \\ 3 \end{array}\right], \left[\begin{array}{r} 4 \\ -2 \\ 1 \end{array}\right]$.

Exercise 4.9.7 *Is* $\vec{u} \times (\vec{v} \times \vec{w}) = (\vec{u} \times \vec{v}) \times \vec{w}$? *What is the meaning of* $\vec{u} \times \vec{v} \times \vec{w}$? *Explain.*
Hint: *Try* $\left(\vec{i} \times \vec{j}\right) \times \vec{k}$.

Exercise 4.9.8 *Verify directly that the coordinate description of the cross product,* $\vec{u} \times \vec{v}$ *has the property that it is perpendicular to both* $\vec{u}$ *and* $\vec{v}$. *Then show by direct computation that this coordinate description satisfies*

$$\begin{aligned} \|\vec{u} \times \vec{v}\|^2 &= \|\vec{u}\|^2\|\vec{v}\|^2 - (\vec{u} \bullet \vec{v})^2 \\ &= \|\vec{u}\|^2\|\vec{v}\|^2 \left(1 - \cos^2(\theta)\right) \end{aligned}$$

where θ *is the angle included between the two vectors. Explain why* $\|\vec{u} \times \vec{v}\|$ *has the correct magnitude.*

Exercise 4.9.9 *Suppose* A *is a* 3×3 *skew symmetric matrix such that* $A^T = -A$. *Show there exists a vector* $\vec{\Omega}$ *such that for all* $\vec{u} \in \mathbb{R}^3$

$$A\vec{u} = \vec{\Omega} \times \vec{u}$$

Hint: *Explain why, since* A *is skew symmetric it is of the form*

$$A = \left[\begin{array}{ccc} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{array}\right]$$

where the ω_i *are numbers. Then consider* $\omega_1\vec{i} + \omega_2\vec{j} + \omega_3\vec{k}$.

Exercise 4.9.10 *Find the volume of the parallelepiped determined by the vectors* $\begin{bmatrix} 1 \\ -7 \\ -5 \end{bmatrix}$, $\begin{bmatrix} 1 \\ -2 \\ -6 \end{bmatrix}$, *and* $\begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}$.

Exercise 4.9.11 *Suppose* $\vec{u}, \vec{v}$, *and* $\vec{w}$ *are three vectors whose components are all integers. Can you conclude the volume of the parallelepiped determined from these three vectors will always be an integer?*

Exercise 4.9.12 *What does it mean geometrically if the box product of three vectors gives zero?*

Exercise 4.9.13 *Using Problem 4.9.12, find an equation of a plane containing the two position vectors,* $\vec{p}$ *and* $\vec{q}$ *and the point* 0. ***Hint:*** *If* (x, y, z) *is a point on this plane, the volume of the parallelepiped determined by* (x, y, z) *and the vectors* $\vec{p}, \vec{q}$ *equals 0.*

Exercise 4.9.14 *Using the notion of the box product yielding either plus or minus the volume of the parallelepiped determined by the given three vectors, show that*

$$(\vec{u} \times \vec{v}) \bullet \vec{w} = \vec{u} \bullet (\vec{v} \times \vec{w})$$

In other words, the dot and the cross can be switched as long as the order of the vectors remains the same. ***Hint:*** *There are two ways to do this, by the coordinate description of the dot and cross product and by geometric reasoning.*

Exercise 4.9.15 *Simplify* $(\vec{u} \times \vec{v}) \bullet (\vec{v} \times \vec{w}) \times (\vec{w} \times \vec{z})$.

Exercise 4.9.16 *Simplify* $\|\vec{u} \times \vec{v}\|^2 + (\vec{u} \bullet \vec{v})^2 - \|\vec{u}\|^2 \|\vec{v}\|^2$.

Exercise 4.9.17 *For* $\vec{u}, \vec{v}, \vec{w}$ *functions of* t, *prove the following product rules:*

$$\begin{aligned} (\vec{u} \times \vec{v})' &= \vec{u}' \times \vec{v} + \vec{u} \times \vec{v}' \\ (\vec{u} \bullet \vec{v})' &= \vec{u}' \bullet \vec{v} + \vec{u} \bullet \vec{v}' \end{aligned}$$

4.10 Spanning, Linear Independence and Basis in $\mathbb{R}^n$

Outcomes

A. *Determine the span of a set of vectors, and determine if a vector is contained in a specified span.*

B. *Determine if a set of vectors is linearly independent.*

C. *Understand the concepts of subspace, basis, and dimension.*

D. *Find the row space, column space, and null space of a matrix.*

By generating all linear combinations of a set of vectors one can obtain various subsets of $\mathbb{R}^n$ which we call subspaces. For example what set of vectors in $\mathbb{R}^3$ generate the XY-plane? What is the smallest such set of vectors can you find? The tools of spanning, linear independence and basis are exactly what is needed to answer these and similar questions and are the focus of this section. The following definition is essential.

Definition 4.58: Subset

Let U and W be sets of vectors in $\mathbb{R}^n$. If all vectors in U are also in W, we say that U is a ***subset*** *of W, denoted*

$$U \subseteq W$$

4.10.1. Spanning Set of Vectors

We begin this section with a definition.

Definition 4.59: Span of a Set of Vectors

The collection of all linear combinations of a set of vectors $\{\vec{u}_1, \cdots, \vec{u}_k\}$ in $\mathbb{R}^n$ is known as the span of these vectors and is written as span$\{\vec{u}_1, \cdots, \vec{u}_k\}$.

Consider the following example.

Example 4.60: Span of Vectors

Describe the span of the vectors $\vec{u} = \left[\begin{array}{ccc} 1 & 1 & 0 \end{array} \right]^T$ and $\vec{v} = \left[\begin{array}{ccc} 3 & 2 & 0 \end{array} \right]^T \in \mathbb{R}^3$.

Solution. You can see that any linear combination of the vectors $\vec{u}$ and $\vec{v}$ yields a vector of the form $\left[\begin{array}{ccc} x & y & 0 \end{array} \right]^T$ in the XY-plane.

Moreover every vector in the XY-plane is in fact such a linear combination of the vectors $\vec{u}$ and $\vec{v}$. That's because

$$\left[\begin{array}{c} x \\ y \\ 0 \end{array} \right] = (-2x+3y) \left[\begin{array}{c} 1 \\ 1 \\ 0 \end{array} \right] + (x-y) \left[\begin{array}{c} 3 \\ 2 \\ 0 \end{array} \right]$$

Thus $\mathrm{span}\{\vec{u}, \vec{v}\}$ is precisely the XY-plane. ◻

You can convince yourself that no single vector can span the XY-plane. In fact, take a moment to consider what is meant by the span of a single vector.

However you can make the set larger if you wish. For example consider the larger set of vectors $\{\vec{u}, \vec{v}, \vec{w}\}$ where $\vec{w} = \left[\begin{array}{ccc} 4 & 5 & 0 \end{array} \right]^T$. Since the first two vectors already span the entire XY-plane, the span is once again precisely the XY-plane and nothing has been gained. Of course if you add a new vector such as $\vec{w} = \left[\begin{array}{ccc} 0 & 0 & 1 \end{array} \right]^T$ then it does span a different space. What is the span of $\vec{u}, \vec{v}, \vec{w}$ in this case?

The distinction between the sets $\{\vec{u}, \vec{v}\}$ and $\{\vec{u}, \vec{v}, \vec{w}\}$ will be made using the concept of linear independence.

Consider the vectors $\vec{u}, \vec{v}$, and $\vec{w}$ discussed above. In the next example, we will show how to formally demonstrate that $\vec{w}$ is in the span of $\vec{u}$ and $\vec{v}$.

Example 4.61: Vector in a Span

Let $\vec{u} = \left[\begin{array}{ccc} 1 & 1 & 0 \end{array} \right]^T$ *and* $\vec{v} = \left[\begin{array}{ccc} 3 & 2 & 0 \end{array} \right]^T \in \mathbb{R}^3$. *Show that* $\vec{w} = \left[\begin{array}{ccc} 4 & 5 & 0 \end{array} \right]^T$ *is in span* $\{\vec{u}, \vec{v}\}$.

Solution. For a vector to be in $\mathrm{span}\,\{\vec{u}, \vec{v}\}$, it must be a linear combination of these vectors. If $\vec{w} \in \mathrm{span}\,\{\vec{u}, \vec{v}\}$, we must be able to find scalars a, b such that

$$\vec{w} = a\vec{u} + b\vec{v}$$

We proceed as follows.

$$\left[\begin{array}{c} 4 \\ 5 \\ 0 \end{array} \right] = a \left[\begin{array}{c} 1 \\ 1 \\ 0 \end{array} \right] + b \left[\begin{array}{c} 3 \\ 2 \\ 0 \end{array} \right]$$

This is equivalent to the following system of equations

$$\begin{array}{rcl} a + 3b & = & 4 \\ a + 2b & = & 5 \end{array}$$

We solving this system the usual way, constructing the augmented matrix and row reducing to find the reduced row-echelon form.

$$\left[\begin{array}{cc|c} 1 & 3 & 4 \\ 1 & 2 & 5 \end{array} \right] \rightarrow \cdots \rightarrow \left[\begin{array}{cc|r} 1 & 0 & 7 \\ 0 & 1 & -1 \end{array} \right]$$

The solution is $a = 7, b = -1$. This means that

$$\vec{w} = 7\vec{u} - \vec{v}$$

Therefore we can say that $\vec{w}$ is in span $\{\vec{u}, \vec{v}\}$. □

4.10.2. Linearly Independent Set of Vectors

Together with the notion of spanning, linear independence is a very important property of a set of vectors.

Definition 4.62: Linearly Independent Set of Vectors

*A set of non-zero vectors $\{\vec{u}_1, \cdots, \vec{u}_k\}$ in $\mathbb{R}^n$ is said to be **linearly independent** if no vector in that set is in the span of the other vectors of that set.*

Here is an example.

Example 4.63: Linearly Independent Vectors

Consider the vectors $\vec{u} = \left[\begin{array}{ccc} 1 & 1 & 0 \end{array}\right]^T$, $\vec{v} = \left[\begin{array}{ccc} 3 & 2 & 0 \end{array}\right]^T$, and $\vec{w} = \left[\begin{array}{ccc} 4 & 5 & 0 \end{array}\right]^T$ in $\mathbb{R}^3$. Verify whether the set $\{\vec{u}, \vec{v}, \vec{w}\}$ is linearly independent.

Solution. We already verified in Example 4.115 that $\vec{w} \in \text{span}\{\vec{u}, \vec{v}\}$. Therefore the set $\{\vec{u}, \vec{v}, \vec{w}\}$ is **not** linearly independent. In this case we say it is linearly dependent. □

In terms of spanning, a set of vectors is linearly independent if it does not contain unnecessary vectors. In the previous example you can see that the vector $\vec{w}$ does not help to span any new vector not already in the span of the other two vectors. However you can verify that the set $\{\vec{u}, \vec{v}\}$ is linearly independent, since both are required to span the XY-plane.

Consider the following important theorem.

Theorem 4.64: Linear Independence as a Linear Combination

The collection of vectors $\{\vec{u}_1, \cdots, \vec{u}_k\}$ in $\mathbb{R}^n$ is linearly independent if and only if whenever

$$\sum_{i=1}^{k} a_i \vec{u}_i = \vec{0}$$

it follows that each $a_i = 0$.
In other words, $\{\vec{u}_1, \cdots, \vec{u}_k\}$ in $\mathbb{R}^n$ is linearly independent exactly when the system of linear equations $AX = 0$ has only the trivial solution, where A is the $n \times k$ matrix having these vectors as columns.

Proof. Suppose first $\{\vec{u}_1, \cdots, \vec{u}_k\}$ is linearly independent. Then by Definition 4.116 none of the vectors is a linear combination of the others. Now suppose for the sake of a contradiction that

$$\sum_{i=1}^{k} a_i \vec{u}_i = \vec{0}$$

and not all the $a_i = 0$. Then pick any a_i which is not zero and divide this equation by it. Solving for $\vec{u}_i$ in terms of the other $\vec{u}_j$ contradicts the fact that none of the $\vec{u}_i$ equals a linear combination of the others. Therefore if the set of vectors is linearly independent and a linear combination of these vectors equals the zero vector, then all the coefficients must equal zero.

Now suppose that whenever a linear combination of the vectors equals the zero vector, then all coefficients equal zero. We want to show that the vectors are linearly independent. If $\vec{u}_i$ is a linear combination of the other vectors in the list, then you could obtain an equation of the form

$$\vec{u}_i = \sum_{j \neq i} a_j \vec{u}_j$$

and so we could write

$$\vec{0} = \sum_{j \neq i} a_j \vec{u}_j + (-1)\, \vec{u}_i$$

which yields a linear combination of the vectors equaling the zero vector, but contradicting the condition that all coefficients must equal 0.

Finally observe that the expression $\sum_{i=1}^{k} a_i \vec{u}_i = \vec{0}$ can be written as the system of linear equations $AX = 0$ where A is the $n \times k$ matrix having these vectors as columns. This explains the last sentence of the theorem. □

The last sentence of this theorem is useful as it allows us to use the reduced row-echelon form of a matrix to determine if a set of vectors is linearly independent. Let the vectors be columns of a matrix A. Find the reduced row-echelon form of A. If each column has a leading one, then it follows that the vectors are linearly independent.

Sometimes we refer to the condition regarding sums as follows: The set of vectors, $\{\vec{u}_1, \cdots, \vec{u}_k\}$ is linearly independent if and only if there is no nontrivial linear combination which equals the zero vector. A nontrivial linear combination is one in which not all the scalars equal zero. Similarly, a trivial linear combination is one in which all scalars equal zero.

We can say that a set of vectors is **linearly dependent** if it is not linearly independent, and hence if at least one vector is a linear combination of the others.

Here is a detailed example in $\mathbb{R}^4$.

Example 4.65: Linear Independence

Determine whether the set of vectors given by

$$\left\{ \left[\begin{array}{c} 1 \\ 2 \\ 3 \\ 0 \end{array} \right], \left[\begin{array}{c} 2 \\ 1 \\ 0 \\ 1 \end{array} \right], \left[\begin{array}{c} 0 \\ 1 \\ 1 \\ 2 \end{array} \right], \left[\begin{array}{c} 3 \\ 2 \\ 2 \\ 0 \end{array} \right] \right\}$$

is linearly independent. If it is linearly dependent, express one of the vectors as a linear combination of the others.

Solution. In this case the matrix of the corresponding homogeneous system of linear equations is

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

The reduced row-echelon form is

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

and so every column is a pivot column and the corresponding system $AX = 0$ only has the trivial solution. Therefore, these vectors are linearly independent and there is no way to obtain one of the vectors as a linear combination of the others. ☐

Consider another example.

Example 4.66: Linear Independence

Determine whether the set of vectors given by

$$\left\{ \left[\begin{array}{c} 1 \\ 2 \\ 3 \\ 0 \end{array} \right], \left[\begin{array}{c} 2 \\ 1 \\ 0 \\ 1 \end{array} \right], \left[\begin{array}{c} 0 \\ 1 \\ 1 \\ 2 \end{array} \right], \left[\begin{array}{r} 3 \\ 2 \\ 2 \\ -1 \end{array} \right] \right\}$$

is linearly independent. If it is linearly dependent, express one of the vectors as a linear combination of the others.

Solution. Form the 4×4 matrix A having these vectors as columns:

$$A = \left[\begin{array}{rrrr} 1 & 2 & 0 & 3 \\ 2 & 1 & 1 & 2 \\ 3 & 0 & 1 & 2 \\ 0 & 1 & 2 & -1 \end{array} \right]$$

Then by Theorem 4.64, the given set of vectors is linearly independent exactly if the system $AX = 0$ has only the trivial solution.

The augmented matrix for this system and corresponding reduced row-echelon form are given by

$$\left[\begin{array}{rrrr|r} 1 & 2 & 0 & 3 & 0 \\ 2 & 1 & 1 & 2 & 0 \\ 3 & 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & -1 & 0 \end{array}\right] \rightarrow \cdots \rightarrow \left[\begin{array}{rrrr|r} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Not all the columns of the coefficient matrix are pivot columns and so the vectors are not linearly independent. In this case, we say the vectors are linearly dependent.

It follows that there are infinitely many solutions to $AX = 0$, one of which is

$$\left[\begin{array}{r} 1 \\ 1 \\ -1 \\ -1 \end{array}\right]$$

Therefore we can write

$$1\left[\begin{array}{r} 1 \\ 2 \\ 3 \\ 0 \end{array}\right] + 1\left[\begin{array}{r} 2 \\ 1 \\ 0 \\ 1 \end{array}\right] - 1\left[\begin{array}{r} 0 \\ 1 \\ 1 \\ 2 \end{array}\right] - 1\left[\begin{array}{r} 3 \\ 2 \\ 2 \\ -1 \end{array}\right] = \left[\begin{array}{r} 0 \\ 0 \\ 0 \\ 0 \end{array}\right]$$

This can be rearranged as follows

$$1\left[\begin{array}{r} 1 \\ 2 \\ 3 \\ 0 \end{array}\right] + 1\left[\begin{array}{r} 2 \\ 1 \\ 0 \\ 1 \end{array}\right] - 1\left[\begin{array}{r} 0 \\ 1 \\ 1 \\ 2 \end{array}\right] = \left[\begin{array}{r} 3 \\ 2 \\ 2 \\ -1 \end{array}\right]$$

This gives the last vector as a linear combination of the first three vectors.

Notice that we could rearrange this equation to write any of the four vectors as a linear combination of the other three. ◻

When given a linearly independent set of vectors, we can determine if related sets are linearly independent.

Example 4.67: Related Sets of Vectors

Let $\{\vec{u}, \vec{v}, \vec{w}\}$ be an independent set of $\mathbb{R}^n$. Is $\{\vec{u}+\vec{v}, 2\vec{u}+\vec{w}, \vec{v}-5\vec{w}\}$ linearly independent?

Solution. Suppose $a(\vec{u}+\vec{v}) + b(2\vec{u}+\vec{w}) + c(\vec{v}-5\vec{w}) = \vec{0}_n$ for some $a, b, c \in \mathbb{R}$. Then

$$(a+2b)\vec{u} + (a+c)\vec{v} + (b-5c)\vec{w} = \vec{0}_n.$$

Since $\{\vec{u}, \vec{v}, \vec{w}\}$ is independent,

$$\begin{aligned} a + 2b &= 0 \\ a + c &= 0 \\ b - 5c &= 0 \end{aligned}$$

This system of three equations in three variables has the unique solution $a = b = c = 0$. Therefore, $\{\vec{u} + \vec{v}, 2\vec{u} + \vec{w}, \vec{v} - 5\vec{w}\}$ is independent. ◻

The following corollary follows from the fact that if the augmented matrix of a homogeneous system of linear equations has more columns than rows, the system has infinitely many solutions.

Corollary 4.68: Linear Dependence in $\mathbb{R}^n$

Let $\{\vec{u}_1, \cdots, \vec{u}_k\}$ be a set of vectors in $\mathbb{R}^n$. If $k > n$, then the set is linearly dependent (i.e. NOT linearly independent).

Proof. Form the $n \times k$ matrix A having the vectors $\{\vec{u}_1, \cdots, \vec{u}_k\}$ as its columns and suppose $k > n$. Then A has rank $r \leq n < k$, so the system $AX = 0$ has a nontrivial solution and thus not linearly independent by Theorem 4.64. ◻

Example 4.69: Linear Dependence

Consider the vectors

$$\left\{ \left[\begin{array}{c} 1 \\ 4 \end{array} \right], \left[\begin{array}{c} 2 \\ 3 \end{array} \right], \left[\begin{array}{c} 3 \\ 2 \end{array} \right] \right\}$$

Are these vectors linearly independent?

Solution. This set contains three vectors in $\mathbb{R}^2$. By Corollary 4.68 these vectors are linearly dependent. In fact, we can write

$$(-1) \left[\begin{array}{c} 1 \\ 4 \end{array} \right] + (2) \left[\begin{array}{c} 2 \\ 3 \end{array} \right] = \left[\begin{array}{c} 3 \\ 2 \end{array} \right]$$

showing that this set is linearly dependent. ◻

The third vector in the previous example is in the span of the first two vectors. We could find a way to write this vector as a linear combination of the other two vectors. It turns out that the linear combination which we found is the **only** one, provided that the set is linearly independent.

Theorem 4.70: Unique Linear Combination

Let $U \subseteq \mathbb{R}^n$ be an independent set. Then any vector $\vec{x} \in span(U)$ can be written uniquely as a linear combination of vectors of U.

Proof. To prove this theorem, we will show that two linear combinations of vectors in U that equal $\vec{x}$ must be the same. Let $U = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$. Suppose that there is a vector $\vec{x} \in \mathrm{span}(U)$ such that

$$\begin{aligned}\vec{x} &= s_1\vec{u}_1 + s_2\vec{u}_2 + \cdots + s_k\vec{u}_k, \text{ for some } s_1, s_2, \ldots, s_k \in \mathbb{R}, \text{ and} \\ \vec{x} &= t_1\vec{u}_1 + t_2\vec{u}_2 + \cdots + t_k\vec{u}_k, \text{ for some } t_1, t_2, \ldots, t_k \in \mathbb{R}.\end{aligned}$$

Then $\vec{0}_n = \vec{x} - \vec{x} = (s_1 - t_1)\vec{u}_1 + (s_2 - t_2)\vec{u}_2 + \cdots + (s_k - t_k)\vec{u}_k$.

Since U is independent, the only linear combination that vanishes is the trivial one, so $s_i - t_i = 0$ for all i, $1 \leq i \leq k$.

Therefore, $s_i = t_i$ for all i, $1 \leq i \leq k$, and the representation is unique. □

Suppose that $\vec{u}, \vec{v}$ and $\vec{w}$ are nonzero vectors in $\mathbb{R}^3$, and that $\{\vec{v}, \vec{w}\}$ is independent. Consider the set $\{\vec{u}, \vec{v}, \vec{w}\}$. When can we know that this set is independent? It turns out that this follows exactly when $\vec{u} \not\in \mathrm{span}\{\vec{v}, \vec{w}\}$.

Example 4.71

Suppose that $\vec{u}, \vec{v}$ and $\vec{w}$ are nonzero vectors in $\mathbb{R}^3$, and that $\{\vec{v}, \vec{w}\}$ is independent. Prove that $\{\vec{u}, \vec{v}, \vec{w}\}$ is independent if and only if $\vec{u} \not\in$ span$\{\vec{v}, \vec{w}\}$.

Solution. If $\vec{u} \in \mathrm{span}\{\vec{v}, \vec{w}\}$, then there exist $a, b \in \mathbb{R}$ so that $\vec{u} = a\vec{v} + b\vec{w}$. This implies that $\vec{u} - a\vec{v} - b\vec{w} = \vec{0}_3$, so $\vec{u} - a\vec{v} - b\vec{w}$ is a nontrivial linear combination of $\{\vec{u}, \vec{v}, \vec{w}\}$ that vanishes, and thus $\{\vec{u}, \vec{v}, \vec{w}\}$ is dependent.

Now suppose that $\vec{u} \not\in \mathrm{span}\{\vec{v}, \vec{w}\}$, and suppose that there exist $a, b, c \in \mathbb{R}$ such that $a\vec{u} + b\vec{v} + c\vec{w} = \vec{0}_3$. If $a \neq 0$, then $\vec{u} = -\frac{b}{a}\vec{v} - \frac{c}{a}\vec{w}$, and $\vec{u} \in \mathrm{span}\{\vec{v}, \vec{w}\}$, a contradiction. Therefore, $a = 0$, implying that $b\vec{v} + c\vec{w} = \vec{0}_3$. Since $\{\vec{v}, \vec{w}\}$ is independent, $b = c = 0$, and thus $a = b = c = 0$, i.e., the only linear combination of $\vec{u}, \vec{v}$ and $\vec{w}$ that vanishes is the trivial one.

Therefore, $\{\vec{u}, \vec{v}, \vec{w}\}$ is independent. □

Consider the following useful theorem.

Theorem 4.72: Invertible Matrices

Let A be an invertible $n \times n$ matrix. Then the columns of A are independent and span $\mathbb{R}^n$. Similarly, the rows of A are independent and span the set of all $1 \times n$ vectors.

This theorem also allows us to determine if a matrix is invertible. If an $n \times n$ matrix A has columns which are independent, or span $\mathbb{R}^n$, then it follows that A is invertible. If it has rows that are independent, or span the set of all $1 \times n$ vectors, then A is invertible.

4.10.3. Subspaces and Basis

The goal of this section is to develop an understanding of a subspace of $\mathbb{R}^n$. Before a precise definition is considered, we first examine the subspace test given below.

Theorem 4.73: Subspace Test

A subset V of $\mathbb{R}^n$ is a subspace of $\mathbb{R}^n$ if

1. *the zero vector of $\mathbb{R}^n$, $\vec{0}_n$, is in V;*
2. *V is closed under addition, i.e., for all $\vec{u}, \vec{w} \in V$, $\vec{u} + \vec{w} \in V$;*
3. *V is closed under scalar multiplication, i.e., for all $\vec{u} \in V$ and $k \in \mathbb{R}$, $k\vec{u} \in V$.*

This test allows us to determine if a given set is a subspace of $\mathbb{R}^n$. Notice that the subset $V = \left\{\vec{0}\right\}$ is a subspace of $\mathbb{R}^n$ (called the zero subspace), as is $\mathbb{R}^n$ itself. A subspace which is not the zero subspace of $\mathbb{R}^n$ is referred to as a proper subspace.

A subspace is simply a set of vectors with the property that linear combinations of these vectors remain in the set. Geometrically in $\mathbb{R}^3$, it turns out that a subspace can be represented by either the origin as a single point, lines and planes which contain the origin, or the entire space $\mathbb{R}^3$.

Consider the following example of a line in $\mathbb{R}^3$.

Example 4.74: Subspace of $\mathbb{R}^3$

In $\mathbb{R}^3$, the line L through the origin that is parallel to the vector $\vec{d} = \begin{bmatrix} -5 \\ 1 \\ -4 \end{bmatrix}$ has (vector) equation $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = t \begin{bmatrix} -5 \\ 1 \\ -4 \end{bmatrix}, t \in \mathbb{R}$, so

$$L = \left\{ t\vec{d} \mid t \in \mathbb{R} \right\}.$$

Then L is a subspace of $\mathbb{R}^3$.

Solution. Using the subspace test given above we can verify that L is a subspace of $\mathbb{R}^3$.

- First: $\vec{0}_3 \in L$ since $0\vec{d} = \vec{0}_3$.
- Suppose $\vec{u}, \vec{v} \in L$. Then by definition, $\vec{u} = s\vec{d}$ and $\vec{v} = t\vec{d}$, for some $s, t \in \mathbb{R}$. Thus

$$\vec{u} + \vec{v} = s\vec{d} + t\vec{d} = (s+t)\vec{d}.$$

  Since $s + t \in \mathbb{R}$, $\vec{u} + \vec{v} \in L$; i.e., L is closed under addition.

- Suppose $\vec{u} \in L$ and $k \in \mathbb{R}$ (k is a scalar). Then $\vec{u} = t\vec{d}$, for some $t \in \mathbb{R}$, so

$$k\vec{u} = k(t\vec{d}) = (kt)\vec{d}.$$

  Since $kt \in \mathbb{R}$, $k\vec{u} \in L$; i.e., L is closed under scalar multiplication.

Since L satisfies all conditions of the subspace test, it follows that L is a subspace. ☐

Note that there is nothing special about the vector $\vec{d}$ used in this example; the same proof works for any nonzero vector $\vec{d} \in \mathbb{R}^3$, so any line through the origin is a subspace of $\mathbb{R}^3$.

We are now prepared to examine the precise definition of a subspace as follows.

Definition 4.75: Subspace

Let V be a nonempty collection of vectors in $\mathbb{R}^n$. Then V is called a subspace if whenever a and b are scalars and $\vec{u}$ and $\vec{v}$ are vectors in V, the linear combination $a\vec{u} + b\vec{v}$ is also in V.

More generally this means that a subspace contains the span of any finite collection vectors in that subspace. It turns out that in $\mathbb{R}^n$, a subspace is exactly the span of finitely many of its vectors.

Theorem 4.76: Subspaces are Spans

Let V be a nonempty collection of vectors in $\mathbb{R}^n$. Then V is a subspace of $\mathbb{R}^n$ if and only if there exist vectors $\{\vec{u}_1, \cdots, \vec{u}_k\}$ in V such that

$$V = \text{span}\{\vec{u}_1, \cdots, \vec{u}_k\}$$

Furthermore, let W be another subspace of $\mathbb{R}^n$ and suppose $\{\vec{u}_1, \cdots, \vec{u}_k\} \in W$. Then it follows that V is a subset of W.

Note that since W is arbitrary, the statement that $V \subseteq W$ means that any other subspace of $\mathbb{R}^n$ that contains these vectors will also contain V.

Proof. We first show that if V is a subspace, then it can be written as $V = \text{span}\{\vec{u}_1, \cdots, \vec{u}_k\}$. Pick a vector $\vec{u}_1$ in V. If $V = \text{span}\{\vec{u}_1\}$, then you have found your list of vectors and are done. If $V \neq \text{span}\{\vec{u}_1\}$, then there exists $\vec{u}_2$ a vector of V which is not in $\text{span}\{\vec{u}_1\}$. Consider $\text{span}\{\vec{u}_1, \vec{u}_2\}$. If $V = \text{span}\{\vec{u}_1, \vec{u}_2\}$, we are done. Otherwise, pick $\vec{u}_3$ not in $\text{span}\{\vec{u}_1, \vec{u}_2\}$. Continue this way. Note that since V is a subspace, these spans are each contained in V. The process must stop with $\vec{u}_k$ for some $k \leq n$ by Corollary 4.68, and thus $V = \text{span}\{\vec{u}_1, \cdots, \vec{u}_k\}$.

Now suppose $V = \text{span}\{\vec{u}_1, \cdots, \vec{u}_k\}$, we must show this is a subspace. So let $\sum_{i=1}^{k} c_i\vec{u}_i$ and $\sum_{i=1}^{k} d_i\vec{u}_i$ be two vectors in V, and let a and b be two scalars. Then

$$a\sum_{i=1}^{k} c_i\vec{u}_i + b\sum_{i=1}^{k} d_i\vec{u}_i = \sum_{i=1}^{k} (ac_i + bd_i)\,\vec{u}_i$$

which is one of the vectors in span $\{\vec{u}_1, \cdots, \vec{u}_k\}$ and is therefore contained in V. This shows that span $\{\vec{u}_1, \cdots, \vec{u}_k\}$ has the properties of a subspace.

To prove that $V \subseteq W$, we prove that if $\vec{u}_i \in V$, then $\vec{u}_i \in W$.

Suppose $\vec{u} \in V$. Then $\vec{u} = a_1\vec{u}_1 + a_2\vec{u}_2 + \cdots + a_k\vec{u}_k$ for some $a_i \in \mathbb{R}$, $1 \leq i \leq k$. Since W contain each $\vec{u}_i$ and W is a vector space, it follows that $a_1\vec{u}_1 + a_2\vec{u}_2 + \cdots + a_k\vec{u}_k \in W$.

$\square$

Since the vectors $\vec{u}_i$ we constructed in the proof above are not in the span of the previous vectors (by definition), they must be linearly independent and thus we obtain the following corollary.

Corollary 4.77: Subspaces are Spans of Independent Vectors

If V is a subspace of $\mathbb{R}^n$, then there exist linearly independent vectors $\{\vec{u}_1, \cdots, \vec{u}_k\}$ in V such that $V =$ span $\{\vec{u}_1, \cdots, \vec{u}_k\}$.

In summary, subspaces of $\mathbb{R}^n$ consist of spans of finite, linearly independent collections of vectors of $\mathbb{R}^n$. Such a collection of vectors is called a basis.

Definition 4.78: Basis of a Subspace

*Let V be a subspace of $\mathbb{R}^n$. Then $\{\vec{u}_1, \cdots, \vec{u}_k\}$ is a **basis** for V if the following two conditions hold.*

1. *span $\{\vec{u}_1, \cdots, \vec{u}_k\} = V$*
2. *$\{\vec{u}_1, \cdots, \vec{u}_k\}$ is linearly independent*

*Note the plural of basis is **bases**.*

The following is a simple but very useful example of a basis, called the standard basis.

Definition 4.79: Standard Basis of $\mathbb{R}^n$

Let $\vec{e}_i$ be the vector in $\mathbb{R}^n$ which has a 1 in the i^{th} entry and zeros elsewhere, that is the i^{th} column of the identity matrix. Then the collection $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$ is a basis for $\mathbb{R}^n$ and is called the standard basis of $\mathbb{R}^n$.

The main theorem about bases is not only they exist, but that they must be of the same size. To show this, we will need the the following fundamental result, called the Exchange Theorem.

Theorem 4.80: Exchange Theorem

Suppose $\{\vec{u}_1, \cdots, \vec{u}_r\}$ is a linearly independent set of vectors in $\mathbb{R}^n$, and each $\vec{u}_k$ is contained in span $\{\vec{v}_1, \cdots, \vec{v}_s\}$ Then $s \geq r$.
In words, spanning sets have at least as many vectors as linearly independent sets.

Proof. Since each $\vec{u}_j$ is in span $\{\vec{v}_1, \cdots, \vec{v}_s\}$, there exist scalars a_{ij} such that

$$\vec{u}_j = \sum_{i=1}^{s} a_{ij}\vec{v}_i$$

Suppose for a contradiction that $s < r$. Then the matrix $A = [a_{ij}]$ has fewer rows, s than columns, r. Then the system $AX = 0$ has a non trivial solution $\vec{d}$, that is there is a $\vec{d} \neq \vec{0}$ such that $A\vec{d} = \vec{0}$. In other words,

$$\sum_{j=1}^{r} a_{ij}d_j = 0, \ i = 1, 2, \cdots, s$$

Therefore,

$$\begin{aligned}\sum_{j=1}^{r} d_j\vec{u}_j &= \sum_{j=1}^{r} d_j \sum_{i=1}^{s} a_{ij}\vec{v}_i \\ &= \sum_{i=1}^{s}\left(\sum_{j=1}^{r} a_{ij}d_j\right)\vec{v}_i = \sum_{i=1}^{s} 0\vec{v}_i = 0\end{aligned}$$

which contradicts the assumption that $\{\vec{u}_1, \cdots, \vec{u}_r\}$ is linearly independent, because not all the d_j are zero. Thus this contradiction indicates that $s \geq r$. ☐

We are now ready to show that any two bases are of the same size.

Theorem 4.81: Bases of $\mathbb{R}^n$ are of the Same Size

Let V be a subspace of $\mathbb{R}^n$ with two bases B_1 and B_2. Suppose B_1 contains s vectors and B_2 contains r vectors. Then $s = r$.

Proof. This follows right away from Theorem 4.80. Indeed observe that $B_1 = \{\vec{u}_1, \cdots, \vec{u}_s\}$ is a spanning set for V while $B_2 = \{\vec{v}_1, \cdots, \vec{v}_r\}$ is linearly independent, so $s \geq r$. Similarly $B_2 = \{\vec{v}_1, \cdots, \vec{v}_r\}$ is a spanning set for V while $B_1 = \{\vec{u}_1, \cdots, \vec{u}_s\}$ is linearly independent, so $r \geq s$. ☐

The following definition can now be stated.

Definition 4.82: Dimension of a Subspace

*Let V be a subspace of $\mathbb{R}^n$. Then the **dimension** of V, written $\dim(V)$ is defined to be the number of vectors in a basis.*

The next result follows.

Corollary 4.83: Dimension of $\mathbb{R}^n$

The dimension of $\mathbb{R}^n$ is n.

Proof. You only need to exhibit a basis for $\mathbb{R}^n$ which has n vectors. Such a basis is the standard basis $\{\vec{e}_1, \cdots, \vec{e}_n\}$. □

Consider the following example.

Example 4.84: Basis of Subspace

Let

$$V = \left\{ \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \in \mathbb{R}^4 \ : \ a - b = d - c \right\}.$$

Show that V is a subspace of $\mathbb{R}^4$, find a basis of V, and find $\dim(V)$.

Solution. The condition $a - b = d - c$ is equivalent to the condition $a = b - c + d$, so we may write

$$V = \left\{ \begin{bmatrix} b - c + d \\ b \\ c \\ d \end{bmatrix} \ : \ b, c, d \in \mathbb{R} \right\} = \left\{ b \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + c \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + d \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \ : \ b, c, d \in \mathbb{R} \right\}$$

This shows that V is a subspace of $\mathbb{R}^4$, since $V = \text{span}\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ where

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \vec{u}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Furthermore,

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is linearly independent, as can be seen by taking the reduced row-echelon form of the matrix whose columns are $\vec{u}_1, \vec{u}_2$ and $\vec{u}_3$.

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Since every column of the reduced row-echelon form matrix has a leading one, the columns are linearly independent.

Therefore $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ is linearly independent and spans V, so is a basis of V. Hence V has dimension three. $\square$

We continue by stating further properties of a set of vectors in $\mathbb{R}^n$.

Corollary 4.85: Linearly Independent and Spanning Sets in $\mathbb{R}^n$

The following properties hold in $\mathbb{R}^n$:

- *Suppose $\{\vec{u}_1, \cdots, \vec{u}_n\}$ is linearly independent. Then $\{\vec{u}_1, \cdots, \vec{u}_n\}$ is a basis for $\mathbb{R}^n$.*
- *Suppose $\{\vec{u}_1, \cdots, \vec{u}_m\}$ spans $\mathbb{R}^n$. Then $m \geq n$.*
- *If $\{\vec{u}_1, \cdots, \vec{u}_n\}$ spans $\mathbb{R}^n$, then $\{\vec{u}_1, \cdots, \vec{u}_n\}$ is linearly independent.*

Proof. Assume first that $\{\vec{u}_1, \cdots, \vec{u}_n\}$ is linearly independent, and we need to show that this set spans $\mathbb{R}^n$. To do so, let $\vec{v}$ be a vector of $\mathbb{R}^n$, and we need to write $\vec{v}$ as a linear combination of $\vec{u}_i$'s. Consider the matrix A having the vectors $\vec{u}_i$ as columns:

$$A = \left[\begin{array}{ccc} \vec{u}_1 & \cdots & \vec{u}_n \end{array} \right]$$

By linear independence of the $\vec{u}_i$'s, the reduced row-echelon form of A is the identity matrix. Therefore the system $A\vec{x} = \vec{v}$ has a (unique) solution, so $\vec{v}$ is a linear combination of the $\vec{u}_i$'s.

To establish the second claim, suppose that $m < n$. Then letting $\vec{u}_{i_1}, \cdots, \vec{u}_{i_k}$ be the pivot columns of the matrix

$$\left[\begin{array}{ccc} \vec{u}_1 & \cdots & \vec{u}_m \end{array} \right]$$

it follows $k \leq m < n$ and these k pivot columns would be a basis for $\mathbb{R}^n$ having fewer than n vectors, contrary to Corollary 4.83.

Finally consider the third claim. If $\{\vec{u}_1, \cdots, \vec{u}_n\}$ is not linearly independent, then replace this list with $\{\vec{u}_{i_1}, \cdots, \vec{u}_{i_k}\}$ where these are the pivot columns of the matrix

$$\left[\begin{array}{ccc} \vec{u}_1 & \cdots & \vec{u}_n \end{array} \right]$$

Then $\{\vec{u}_{i_1}, \cdots, \vec{u}_{i_k}\}$ spans $\mathbb{R}^n$ and is linearly independent, so it is a basis having less than n vectors again contrary to Corollary 4.83. $\square$

The next theorem follows from the above claim.

Theorem 4.86: Existence of Basis

Let V be a subspace of $\mathbb{R}^n$. Then there exists a basis of V with $\dim(V) \leq n$.

Consider Corollary 4.85 together with Theorem 4.86. Let $\dim(V) = r$. Suppose there exists an independent set of vectors in V. If this set contains r vectors, then it is a basis for V. If it contains less than r vectors, then vectors can be added to the set to create a basis of V. Similarly, any spanning set of V which contains more than r vectors can have vectors removed to create a basis of V.

We illustrate this concept in the next example.

Example 4.87: Extending an Independent Set

Consider the set U given by

$$U = \left\{ \left[\begin{array}{c} a \\ b \\ c \\ d \end{array} \right] \in \mathbb{R}^4 \;\middle|\; a - b = d - c \right\}$$

Then U is a subspace of $\mathbb{R}^4$ and $\dim(U) = 3$.
Then

$$S = \left\{ \left[\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \end{array} \right], \left[\begin{array}{c} 2 \\ 3 \\ 3 \\ 2 \end{array} \right] \right\},$$

is an independent subset of U. Therefore S can be extended to a basis of U.

Solution. To extend S to a basis of U, find a vector in U that is **not** in $\text{span}(S)$.

$$\left[\begin{array}{ccc} 1 & 2 & ? \\ 1 & 3 & ? \\ 1 & 3 & ? \\ 1 & 2 & ? \end{array} \right]$$

$$\left[\begin{array}{rrr} 1 & 2 & 1 \\ 1 & 3 & 0 \\ 1 & 3 & -1 \\ 1 & 2 & 0 \end{array} \right] \rightarrow \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{array} \right]$$

Therefore, S can be extended to the following basis of U:

$$\left\{ \left[\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \end{array} \right], \left[\begin{array}{c} 2 \\ 3 \\ 3 \\ 2 \end{array} \right], \left[\begin{array}{r} 1 \\ 0 \\ -1 \\ 0 \end{array} \right] \right\},$$

$\square$

Next we consider the case of removing vectors from a spanning set to result in a basis.

Theorem 4.88: Finding a Basis from a Span

Let W be a subspace. Also suppose that $W = \text{span}\{\vec{w}_1, \cdots, \vec{w}_m\}$. Then there exists a subset of $\{\vec{w}_1, \cdots, \vec{w}_m\}$ which is a basis for W.

Proof. Let S denote the set of positive integers such that for $k \in S$, there exists a subset of $\{\vec{w}_1, \cdots, \vec{w}_m\}$ consisting of exactly k vectors which is a spanning set for W. Thus $m \in S$. Pick the smallest positive integer in S. Call it k. Then there exists $\{\vec{u}_1, \cdots, \vec{u}_k\} \subseteq \{\vec{w}_1, \cdots, \vec{w}_m\}$ such that span $\{\vec{u}_1, \cdots, \vec{u}_k\} = W$. If

$$\sum_{i=1}^{k} c_i \vec{w}_i = \vec{0}$$

and not all of the $c_i = 0$, then you could pick $c_j \neq 0$, divide by it and solve for $\vec{u}_j$ in terms of the others,

$$\vec{w}_j = \sum_{i \neq j} \left(-\frac{c_i}{c_j} \right) \vec{w}_i$$

Then you could delete $\vec{w}_j$ from the list and have the same span. Any linear combination involving $\vec{w}_j$ would equal one in which $\vec{w}_j$ is replaced with the above sum, showing that it could have been obtained as a linear combination of $\vec{w}_i$ for $i \neq j$. Thus $k-1 \in S$ contrary to the choice of k . Hence each $c_i = 0$ and so $\{\vec{u}_1, \cdots, \vec{u}_k\}$ is a basis for W consisting of vectors of $\{\vec{w}_1, \cdots, \vec{w}_m\}$. ☐

The following example illustrates how to carry out this shrinking process which will obtain a subset of a span of vectors which is linearly independent.

Example 4.89: Subset of a Span

Let W be the subspace

$$\text{span} \left\{ \left[\begin{array}{r} 1 \\ 2 \\ -1 \\ 1 \end{array} \right], \left[\begin{array}{r} 1 \\ 3 \\ -1 \\ 1 \end{array} \right], \left[\begin{array}{r} 8 \\ 19 \\ -8 \\ 8 \end{array} \right], \left[\begin{array}{r} -6 \\ -15 \\ 6 \\ -6 \end{array} \right], \left[\begin{array}{r} 1 \\ 3 \\ 0 \\ 1 \end{array} \right], \left[\begin{array}{r} 1 \\ 5 \\ 0 \\ 1 \end{array} \right] \right\}$$

Find a basis for W which consists of a subset of the given vectors.

Solution. You can use the reduced row-echelon form to accomplish this reduction. Form the matrix which has the given vectors as columns.

$$\left[\begin{array}{rrrrrr} 1 & 1 & 8 & -6 & 1 & 1 \\ 2 & 3 & 19 & -15 & 3 & 5 \\ -1 & -1 & -8 & 6 & 0 & 0 \\ 1 & 1 & 8 & -6 & 1 & 1 \end{array} \right]$$

Then take the reduced row-echelon form

$$\left[\begin{array}{rrrrrr} 1 & 0 & 5 & -3 & 0 & -2 \\ 0 & 1 & 3 & -3 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

It follows that a basis for W is

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Since the first, second, and fifth columns are obviously a basis for the column space of the reduced row-echelon form, the same is true for the matrix having the given vectors as columns. $\square$

Consider the following theorems regarding a subspace contained in another subspace.

Theorem 4.90: Subset of a Subspace

Let V and W be subspaces of $\mathbb{R}^n$, and suppose that $W \subseteq V$. Then $\dim(W) \leq \dim(V)$ with equality when $W = V$.

Theorem 4.91: Extending a Basis

Let W be any non-zero subspace $\mathbb{R}^n$ and let $W \subseteq V$ where V is also a subspace of $\mathbb{R}^n$. Then every basis of W can be extended to a basis for V.

The proof is left as an exercise but proceeds as follows. Begin with a basis for W, $\{\vec{w}_1, \cdots, \vec{w}_s\}$ and add in vectors from V until you obtain a basis for V. Not that the process will stop because the dimension of V is no more than n.

Consider the following example.

Example 4.92: Extending a Basis

Let $V = \mathbb{R}^4$ and let

$$W = span \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Extend this basis of W to a basis of $\mathbb{R}^n$.

Solution. An easy way to do this is to take the reduced row-echelon form of the matrix

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.17}$$

Note how the given vectors were placed as the first two columns and then the matrix was extended in such a way that it is clear that the span of the columns of this matrix yield all of $\mathbb{R}^4$. Now determine the pivot columns. The reduced row-echelon form is

$$\left[\begin{array}{rrrrrr} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{array}\right] \tag{4.18}$$

Therefore the pivot columns are

$$\left[\begin{array}{r} 1 \\ 0 \\ 1 \\ 1 \end{array}\right], \left[\begin{array}{r} 0 \\ 1 \\ 0 \\ 1 \end{array}\right], \left[\begin{array}{r} 1 \\ 0 \\ 0 \\ 0 \end{array}\right], \left[\begin{array}{r} 0 \\ 1 \\ 0 \\ 0 \end{array}\right]$$

and now this is an extension of the given basis for W to a basis for $\mathbb{R}^4$.

Why does this work? The columns of 4.17 obviously span $\mathbb{R}^4$. In fact the span of the first four is the same as the span of all six. $\square$

Consider another example.

Example 4.93: Extending a Basis

Let W be the span of $\left[\begin{array}{r} 1 \\ 0 \\ 1 \\ 0 \end{array}\right]$ *in $\mathbb{R}^4$. Let V consist of the span of the vectors*

$$\left[\begin{array}{r} 1 \\ 0 \\ 1 \\ 0 \end{array}\right], \left[\begin{array}{r} 0 \\ 1 \\ 1 \\ 1 \end{array}\right], \left[\begin{array}{r} 7 \\ -6 \\ 1 \\ -6 \end{array}\right], \left[\begin{array}{r} -5 \\ 7 \\ 2 \\ 7 \end{array}\right], \left[\begin{array}{r} 0 \\ 0 \\ 0 \\ 1 \end{array}\right]$$

Find a basis for V which extends the basis for W.

Solution. Note that the above vectors are not linearly independent, but their span, denoted as V is a subspace which does include the subspace W.

Using the process outlined in the previous example, form the following matrix

$$\left[\begin{array}{rrrrr} 1 & 0 & 7 & -5 & 0 \\ 0 & 1 & -6 & 7 & 0 \\ 1 & 1 & 1 & 2 & 0 \\ 0 & 1 & -6 & 7 & 1 \end{array}\right]$$

Next find its reduced row-echelon form

$$\left[\begin{array}{rrrrr} 1 & 0 & 7 & -5 & 0 \\ 0 & 1 & -6 & 7 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

It follows that a basis for V consists of the first two vectors and the last.

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Thus V is of dimension 3 and it has a basis which extends the basis for W. □

4.10.4. Row Space, Column Space, and Null Space of a Matrix

We begin this section with a new definition.

Definition 4.94: Row and Column Space

*Let A be an $m \times n$ matrix. The **column space** of A, written $col(A)$, is the span of the columns. The **row space** of A, written $row(A)$, is the span of the rows.*

Using the reduced row-echelon form , we can obtain an efficient description of the row and column space of a matrix. Consider the following lemma.

Lemma 4.95: Effect of Row Operations on Row Space

Let A and B be $m \times n$ matrices such that A can be carried to B by elementary row [column] operations. Then $row(A) = row(B)$ [$col(A) = col(B)$].

Proof. We will prove that the above is true for row operations, which can be easily applied to column operations.

Let $\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_m$ denote the rows of A.

- If B is obtained from A by a interchanging two rows of A, then A and B have exactly the same rows, so $\mathrm{row}(B) = \mathrm{row}(A)$.
- Suppose $p \neq 0$, and suppose that for some j, $1 \leq j \leq m$, B is obtained from A by multiplying row j by p. Then

  $$\mathrm{row}(B) = \mathrm{span}\{\vec{r}_1, \ldots, p\vec{r}_j, \ldots, \vec{r}_m\}.$$

  Since

  $$\{\vec{r}_1, \ldots, p\vec{r}_j, \ldots, \vec{r}_m\} \subseteq \mathrm{row}(A),$$

  it follows that $\mathrm{row}(B) \subseteq \mathrm{row}(A)$. Conversely, since

  $$\{\vec{r}_1, \ldots, \vec{r}_m\} \subseteq \mathrm{row}(B),$$

  it follows that $\mathrm{row}(A) \subseteq \mathrm{row}(B)$. Therefore, $\mathrm{row}(B) = \mathrm{row}(A)$.

- Suppose $p \neq 0$, and suppose that for some i and j, $1 \leq i, j \leq m$, B is obtained from A by adding p time row j to row i. Without loss of generality, we may assume $i < j$.

  Then

$$\operatorname{row}(B) = \operatorname{span}\{\vec{r}_1, \ldots, \vec{r}_{i-1}, \vec{r}_i + p\vec{r}_j, \ldots, \vec{r}_j, \ldots, \vec{r}_m\}.$$

  Since

$$\{\vec{r}_1, \ldots, \vec{r}_{i-1}, \vec{r}_i + p\vec{r}_j, \ldots, \vec{r}_m\} \subseteq \operatorname{row}(A),$$

  it follows that $\operatorname{row}(B) \subseteq \operatorname{row}(A)$.

  Conversely, since

$$\{\vec{r}_1, \ldots, \vec{r}_m\} \subseteq \operatorname{row}(B),$$

  it follows that $\operatorname{row}(A) \subseteq \operatorname{row}(B)$. Therefore, $\operatorname{row}(B) = \operatorname{row}(A)$.

□

Consider the following lemma.

Lemma 4.96: Row Space of a Row-Echelon Form Matrix

Let A be an $m \times n$ matrix and let R be its row-echelon form. Then the nonzero rows of R form a basis of row(R), and consequently of row(A).

This lemma suggests that we can examine the row-echelon form of a matrix in order to obtain the row space. Consider now the column space. The column space can be obtained by simply saying that it equals the span of all the columns. However, you can often get the column space as the span of fewer columns than this. A variation of the previous lemma provides a solution. Suppose A is row reduced to its row-echelon form R. Identify the pivot columns of R (columns which have leading ones), and take the corresponding columns of A. It turns out that this forms a basis of $\operatorname{col}(A)$.

Before proceeding to an example of this concept, we revisit the definition of rank.

Definition 4.97: Rank of a Matrix

Previously, we defined rank(A) to be the number of leading entries in the row-echelon form of A. Using an understanding of dimension and row space, we can now define rank as follows:

$$\mathit{rank}(A) = \dim(\mathit{row}(A))$$

Consider the following example.

Example 4.98: Rank, Column and Row Space

Find the rank of the following matrix and describe the column and row spaces.

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

Solution. The reduced row-echelon form of A is

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

Therefore, the rank is 2.

Notice that the first two columns of R are pivot columns. By the discussion following Lemma 4.96, we find the corresponding columns of A, in this case the first two columns. Therefore a basis for $\mathrm{col}(A)$ is given by

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 7 \end{bmatrix} \right\}$$

For example, consider the third column of the original matrix. It can be written as a linear combination of the first two columns of the original matrix as follows.

$$\begin{bmatrix} 1 \\ 6 \\ 8 \end{bmatrix} = -9 \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} + 5 \begin{bmatrix} 2 \\ 3 \\ 7 \end{bmatrix}$$

What about an efficient description of the row space? By Lemma 4.96 we know that the nonzero rows of R create a basis of $\mathrm{row}(A)$. For the above matrix, the row space equals

$$\mathrm{row}(A) = \mathrm{span} \left\{ \begin{bmatrix} 1 & 0 & -9 & 9 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 5 & -3 & 0 \end{bmatrix} \right\}$$

$\square$

Notice that the column space of A is given as the span of columns of the original matrix, while the row space of A is the span of rows of the reduced row-echelon form of A.

Consider another example.

Example 4.99: Rank, Column and Row Space

Find the rank of the following matrix and describe the column and row spaces.

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

Solution. The reduced row-echelon form is

$$\left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & \frac{13}{2} \\ 0 & 1 & 0 & 2 & -\frac{5}{2} \\ 0 & 0 & 1 & -1 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

and so the rank is 3. The row space is given by

$$\mathrm{row}(A) = \mathrm{span}\left\{ \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & \frac{13}{2} \end{array}\right], \left[\begin{array}{ccccc} 0 & 1 & 0 & 2 & -\frac{5}{2} \end{array}\right], \left[\begin{array}{ccccc} 0 & 0 & 1 & -1 & \frac{1}{2} \end{array}\right] \right\}$$

Notice that the first three columns of the reduced row-echelon form are pivot columns. The column space is the span of the first three columns in the **original matrix**,

$$\mathrm{col}(A) = \mathrm{span}\left\{ \left[\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \end{array}\right], \left[\begin{array}{c} 2 \\ 3 \\ 2 \\ 3 \end{array}\right], \left[\begin{array}{c} 1 \\ 6 \\ 1 \\ 2 \end{array}\right] \right\}$$

□

Consider the solution given above for Example 4.99, where the rank of A equals 3. Notice that the row space and the column space each had dimension equal to 3. It turns out that this is not a coincidence, and this essential result is referred to as the Rank Theorem and is given now. Recall that we defined $\mathrm{rank}(A) = \dim(\mathrm{row}(A))$.

Theorem 4.100: Rank Theorem

Let A be an $m \times n$ matrix. Then $\dim(\mathrm{col}(A))$, the dimension of the column space, is equal to the dimension of the row space, $\dim(\mathrm{row}(A))$.

The following statements all follow from the Rank Theorem.

Corollary 4.101: Results of the Rank Theorem

Let A be a matrix. Then the following are true:

1. $\mathrm{rank}(A) = \mathrm{rank}(A^T)$.
2. *For A of size $m \times n$, $\mathrm{rank}(A) \leq m$ and $\mathrm{rank}(A) \leq n$.*
3. *For A of size $n \times n$, A is invertible if and only if $\mathrm{rank}(A) = n$.*
4. *For invertible matrices B and C of appropriate size, $\mathrm{rank}(A) = \mathrm{rank}(BA) = \mathrm{rank}(AC)$.*

Consider the following example.

Example 4.102: Rank of the Transpose

Let

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

Find $\operatorname{rank}(A)$ *and* $\operatorname{rank}(A^T)$.

Solution. To find $\operatorname{rank}(A)$ we first row reduce to find the reduced row-echelon form.

$$A = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \rightarrow \cdots \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Therefore the rank of A is 2. Now consider A^T given by

$$A^T = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}$$

Again we row reduce to find the reduced row-echelon form.

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

You can see that $\operatorname{rank}(A^T) = 2$, the same as $\operatorname{rank}(A)$. ◻

We now define what is meant by the null space of a general $m \times n$ matrix.

Definition 4.103: Null Space, or Kernel, of A

The null space of a matrix A, also referred to as the kernel of A, is defined as follows.

$$\mathit{null}\,(A) = \left\{ \vec{x} : A\vec{x} = \vec{0} \right\}$$

It can also be referred to using the notation $\ker(A)$. Similarly, we can discuss the image of A, denoted by $\operatorname{im}(A)$. The image of A consists of the vectors of $\mathbb{R}^m$ which "get hit" by A. The formal definition is as follows.

Definition 4.104: Image of A

The image of A, written $\mathit{im}\,(A)$ is given by

$$\mathit{im}\,(A) = \left\{ A\vec{x} : \vec{x} \in \mathbb{R}^n \right\}$$

Consider A as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ whose action is given by multiplication. The following diagram displays this scenario.

$$\begin{array}{ccc} \text{null}(A) & & \text{im}(A) \\ \mathbb{R}^n & \xrightarrow{A} & \mathbb{R}^m \end{array}$$

As indicated, $\operatorname{im}(A)$ is a subset of $\mathbb{R}^m$ while $\operatorname{null}(A)$ is a subset of $\mathbb{R}^n$.

It turns out that the null space and image of A are both subspaces. Consider the following example.

Example 4.105: Null Space

Let A be an $m \times n$ matrix. Then the null space of A, null(A) is a subspace of $\mathbb{R}^n$.

Solution.

- Since $A\vec{0}_n = \vec{0}_m$, $\vec{0}_n \in \operatorname{null}(A)$.
- Let $\vec{x}, \vec{y} \in \operatorname{null}(A)$. Then $A\vec{x} = \vec{0}_m$ and $A\vec{y} = \vec{0}_m$, so
$$A(\vec{x}+\vec{y}) = A\vec{x} + A\vec{y} = \vec{0}_m + \vec{0}_m = \vec{0}_m,$$
and thus $\vec{x}+\vec{y} \in \operatorname{null}(A)$.
- Let $\vec{x} \in \operatorname{null}(A)$ and $k \in \mathbb{R}$. Then $A\vec{x} = \vec{0}_m$, so
$$A(k\vec{x}) = k(A\vec{x}) = k\vec{0}_m = \vec{0}_m,$$
and thus $k\vec{x} \in \operatorname{null}(A)$.

Therefore by the subspace test, $\operatorname{null}(A)$ is a subspace of $\mathbb{R}^n$.

$\square$

The proof that $\operatorname{im}(A)$ is a subspace of $\mathbb{R}^m$ is similar and is left as an exercise to the reader.

We now wish to find a way to describe $\operatorname{null}(A)$ for a matrix A. However, finding $\operatorname{null}(A)$ is not new! There is just some new terminology being used, as $\operatorname{null}(A)$ is simply the solution to the system $A\vec{x} = \vec{0}$.

Theorem 4.106: Basis of null(A)

Let A be an $m \times n$ matrix such that rank$(A) = r$. Then the system $A\vec{x} = \vec{0}_m$ has $n - r$ basic solutions, providing a basis of null(A) with $\dim(\text{null}(A)) = n - r$.

Consider the following example.

Example 4.107: Null Space of A

Let
$$A = \left[\begin{array}{rrr} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 2 & 3 & 3 \end{array}\right]$$
Find null(A) and im(A).

Solution. In order to find $\text{null}(A)$, we simply need to solve the equation $A\vec{x} = \vec{0}$. This is the usual procedure of writing the augmented matrix, finding the reduced row-echelon form and then the solution. The augmented matrix and corresponding reduced row-echelon form are

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

The third column is not a pivot column, and therefore the solution will contain a parameter. The solution to the system $A\vec{x} = \vec{0}$ is given by

$$\left[\begin{array}{r} -3t \\ t \\ t \end{array}\right] : t \in \mathbb{R}$$

which can be written as

$$t\left[\begin{array}{r} -3 \\ 1 \\ 1 \end{array}\right] : t \in \mathbb{R}$$

Therefore, the null space of A is all multiples of this vector, which we can write as

$$\text{null}(A) = \text{span}\left\{\left[\begin{array}{r} -3 \\ 1 \\ 1 \end{array}\right]\right\}$$

Finally $\text{im}(A)$ is just $\{A\vec{x} : \vec{x} \in \mathbb{R}^n\}$ and hence consists of the span of all columns of A, that is $\text{im}(A) = \text{col}(A)$.

Notice from the above calculation that that the first two columns of the reduced row-echelon form are pivot columns. Thus the column space is the span of the first two columns in the **original matrix**, and we get

$$\text{im}(A) = \text{col}(A) = \text{span}\left\{\left[\begin{array}{r} 1 \\ 0 \\ 2 \end{array}\right], \left[\begin{array}{r} 2 \\ -1 \\ 3 \end{array}\right]\right\}$$

□

Here is a larger example, but the method is entirely similar.

Example 4.108: Null Space of A

Let

$$A = \left[\begin{array}{rrrrr} 1 & 2 & 1 & 0 & 1 \\ 2 & -1 & 1 & 3 & 0 \\ 3 & 1 & 2 & 3 & 1 \\ 4 & -2 & 2 & 6 & 0 \end{array}\right]$$

Find the null space of A.

Solution. To find the null space, we need to solve the equation $AX = 0$. The augmented matrix and corresponding reduced row-echelon form are given by

$$\left[\begin{array}{rrrrr|r} 1 & 2 & 1 & 0 & 1 & 0 \\ 2 & -1 & 1 & 3 & 0 & 0 \\ 3 & 1 & 2 & 3 & 1 & 0 \\ 4 & -2 & 2 & 6 & 0 & 0 \end{array}\right] \rightarrow \cdots \rightarrow \left[\begin{array}{rrrrr|r} 1 & 0 & \frac{3}{5} & \frac{6}{5} & \frac{1}{5} & 0 \\ 0 & 1 & \frac{1}{5} & -\frac{3}{5} & \frac{2}{5} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

It follows that the first two columns are pivot columns, and the next three correspond to parameters. Therefore, $\mathrm{null}\,(A)$ is given by

$$\left[\begin{array}{c} \left(-\frac{3}{5}\right)s + \left(-\frac{6}{5}\right)t + \left(\frac{1}{5}\right)r \\ \left(-\frac{1}{5}\right)s + \left(\frac{3}{5}\right)t + \left(-\frac{2}{5}\right)r \\ s \\ t \\ r \end{array}\right] : s, t, r \in \mathbb{R}.$$

We write this in the form

$$s\left[\begin{array}{r} -\frac{3}{5} \\ -\frac{1}{5} \\ 1 \\ 0 \\ 0 \end{array}\right] + t\left[\begin{array}{r} -\frac{6}{5} \\ \frac{3}{5} \\ 0 \\ 1 \\ 0 \end{array}\right] + r\left[\begin{array}{r} \frac{1}{5} \\ -\frac{2}{5} \\ 0 \\ 0 \\ 1 \end{array}\right] : s, t, r \in \mathbb{R}.$$

In other words, the null space of this matrix equals the span of the three vectors above. Thus

$$\mathrm{null}\,(A) = \mathrm{span}\left\{\left[\begin{array}{r} -\frac{3}{5} \\ -\frac{1}{5} \\ 1 \\ 0 \\ 0 \end{array}\right], \left[\begin{array}{r} -\frac{6}{5} \\ \frac{3}{5} \\ 0 \\ 1 \\ 0 \end{array}\right], \left[\begin{array}{r} \frac{1}{5} \\ -\frac{2}{5} \\ 0 \\ 0 \\ 1 \end{array}\right]\right\}$$

$\square$

Notice also that the three vectors above are linearly independent and so the dimension of $\mathrm{null}\,(A)$ is 3. The following is true in general, the number of parameters in the solution of $AX = 0$ equals the dimension of the null space. Recall also that the number of leading ones in the reduced row-echelon formequals the number of pivot columns, which is the rank of the matrix, which is the same as the dimension of either the column or row space.

Before we proceed to an important theorem, we first define what is meant by the nullity of a matrix.

Definition 4.109: Nullity

The dimension of the null space of a matrix is called the ***nullity****, denoted* $\dim(\mathit{null}\,(A))$.

From our observation above we can now state an important theorem.

Theorem 4.110: Rank and Nullity

Let A be an $m \times n$ matrix. Then $\mathit{rank}(A) + \dim(\mathit{null}(A)) = n$.

Consider the following example, which we first explored above in Example 4.107

Example 4.111: Rank and Nullity

Let

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

Find $\mathit{rank}(A)$ and $\dim(\mathit{null}(A))$.

Solution. In the above Example 4.107 we determined that the reduced row-echelon form of A is given by

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

Therefore the rank of A is 2. We also determined that the null space of A is given by

$$\text{null}(A) = \text{span}\left\{ \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \right\}$$

Therefore the nullity of A is 1. It follows from Theorem 4.110 that $\text{rank}(A) + \dim(\text{null}(A)) = 2 + 1 = 3$, which is the number of columns of A. ☐

Collecting various results we conclude this section with two related and important theorems.

Theorem 4.112

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

1. *$\mathit{rank}(A) = n$.*
2. *$\mathit{row}(A) = \mathbb{R}^n$, i.e., the rows of A span $\mathbb{R}^n$.*
3. *The columns of A are independent in $\mathbb{R}^m$.*
4. *If $A\vec{x} = \vec{0}_m$ for some $\vec{x} \in \mathbb{R}^n$, then $\vec{x} = \vec{0}_n$.*

Theorem 4.113

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

1. *$\text{rank}(A) = m$.*
2. *$\text{col}(A) = \mathbb{R}^m$, i.e., the columns of A span $\mathbb{R}^m$.*
3. *The rows of A are independent in $\mathbb{R}^n$.*
4. *The system $A\vec{x} = \vec{b}$ is consistent for every $\vec{b} \in \mathbb{R}^m$.*

Exercises

Exercise 4.10.1 *Here are some vectors.*

$$\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ 7 \\ -4 \end{bmatrix}, \begin{bmatrix} 5 \\ 7 \\ -10 \end{bmatrix}, \begin{bmatrix} 12 \\ 17 \\ -24 \end{bmatrix}$$

Describe the span of these vectors as the span of as few vectors as possible.

Exercise 4.10.2 *Here are some vectors.*

$$\begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 12 \\ 29 \\ -24 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ 9 \\ -4 \end{bmatrix}, \begin{bmatrix} 5 \\ 12 \\ -10 \end{bmatrix},$$

Describe the span of these vectors as the span of as few vectors as possible.

Exercise 4.10.3 *Here are some vectors.*

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

Describe the span of these vectors as the span of as few vectors as possible.

Exercise 4.10.4 *Here are some vectors.*

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

Now here is another vector:

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

Is this vector in the span of the first four vectors? If it is, exhibit a linear combination of the first four vectors which equals this vector, using as few vectors as possible in the linear combination.

Exercise 4.10.5 *Here are some vectors.*

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

Now here is another vector:

$$\begin{bmatrix} 2 \\ -3 \\ -4 \end{bmatrix}$$

Is this vector in the span of the first four vectors? If it is, exhibit a linear combination of the first four vectors which equals this vector, using as few vectors as possible in the linear combination.

Exercise 4.10.6 *Here are some vectors.*

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

Now here is another vector:

$$\begin{bmatrix} 1 \\ 9 \\ 1 \end{bmatrix}$$

Is this vector in the span of the first four vectors? If it is, exhibit a linear combination of the first four vectors which equals this vector, using as few vectors as possible in the linear combination.

Exercise 4.10.7 *Here are some vectors.*

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

Now here is another vector:

$$\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

Is this vector in the span of the first four vectors? If it is, exhibit a linear combination of the first four vectors which equals this vector, using as few vectors as possible in the linear combination.

Exercise 4.10.8 *Here are some vectors.*

$$\left[\begin{array}{r} 1 \\ -1 \\ -2 \end{array}\right], \left[\begin{array}{r} 1 \\ 0 \\ -2 \end{array}\right], \left[\begin{array}{r} 1 \\ -5 \\ -2 \end{array}\right], \left[\begin{array}{r} -1 \\ 5 \\ 2 \end{array}\right]$$

Now here is another vector:

$$\left[\begin{array}{r} 1 \\ 1 \\ -1 \end{array}\right]$$

Is this vector in the span of the first four vectors? If it is, exhibit a linear combination of the first four vectors which equals this vector, using as few vectors as possible in the linear combination.

Exercise 4.10.9 *Here are some vectors.*

$$\left[\begin{array}{r} 1 \\ 0 \\ -2 \end{array}\right], \left[\begin{array}{r} 1 \\ 1 \\ -2 \end{array}\right], \left[\begin{array}{r} 2 \\ -2 \\ -3 \end{array}\right], \left[\begin{array}{r} -1 \\ 4 \\ 2 \end{array}\right]$$

Now here is another vector:

$$\left[\begin{array}{r} -1 \\ -4 \\ 2 \end{array}\right]$$

Is this vector in the span of the first four vectors? If it is, exhibit a linear combination of the first four vectors which equals this vector, using as few vectors as possible in the linear combination.

Exercise 4.10.10 *Suppose $\{\vec{x}_1, \cdots, \vec{x}_k\}$ is a set of vectors from $\mathbb{R}^n$. Show that $\vec{0}$ is in* $span\{\vec{x}_1, \cdots, \vec{x}_k\}$.

Exercise 4.10.11 *Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.*

$$\left[\begin{array}{r} 1 \\ 3 \\ -1 \\ 1 \end{array}\right], \left[\begin{array}{r} 1 \\ 4 \\ -1 \\ 1 \end{array}\right], \left[\begin{array}{r} 1 \\ 4 \\ 0 \\ 1 \end{array}\right], \left[\begin{array}{r} 1 \\ 10 \\ 2 \\ 1 \end{array}\right]$$

Exercise 4.10.12 *Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.*

$$\left[\begin{array}{r} -1 \\ -2 \\ 2 \\ 3 \end{array}\right], \left[\begin{array}{r} -3 \\ -4 \\ 3 \\ 3 \end{array}\right], \left[\begin{array}{r} 0 \\ -1 \\ 4 \\ 3 \end{array}\right], \left[\begin{array}{r} 0 \\ -1 \\ 6 \\ 4 \end{array}\right]$$

Exercise 4.10.13 *Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.*

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

Exercise 4.10.14 *Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.*

$$\begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \\ 34 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 7 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 8 \\ 1 \end{bmatrix}$$

Exercise 4.10.15 *Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others.*

$$\begin{bmatrix} 1 \\ 3 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ -10 \\ 3 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 0 \\ 1 \end{bmatrix}$$

Exercise 4.10.16 *Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.*

$$\begin{bmatrix} 1 \\ 3 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ -5 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ -4 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 10 \\ -14 \\ 1 \end{bmatrix}$$

Exercise 4.10.17 *Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.*

$$\begin{bmatrix} 1 \\ 0 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 8 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 7 \\ 34 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 7 \\ 1 \end{bmatrix}$$

Exercise 4.10.18 *Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.*

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

Exercise 4.10.19 *Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others.*

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

Exercise 4.10.20 *Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.*

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

Exercise 4.10.21 *Here are some vectors in $\mathbb{R}^4$.*

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

Thse vectors can't possibly be linearly independent. Tell why. Next obtain a linearly independent subset of these vectors which has the same span as these vectors. In other words, find a basis for the span of these vectors.

Exercise 4.10.22 *Here are some vectors in $\mathbb{R}^4$.*

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

Thse vectors can't possibly be linearly independent. Tell why. Next obtain a linearly independent subset of these vectors which has the same span as these vectors. In other words, find a basis for the span of these vectors.

Exercise 4.10.23 *Here are some vectors in $\mathbb{R}^4$.*

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

Thse vectors can't possibly be linearly independent. Tell why. Next obtain a linearly independent subset of these vectors which has the same span as these vectors. In other words, find a basis for the span of these vectors.

Exercise 4.10.24 *Here are some vectors in* $\mathbb{R}^4$*.*

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

Thse vectors can't possibly be linearly independent. Tell why. Next obtain a linearly independent subset of these vectors which has the same span as these vectors. In other words, find a basis for the span of these vectors.

Exercise 4.10.25 *Here are some vectors in* $\mathbb{R}^4$*.*

$$\begin{bmatrix} 1 \\ 4 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 11 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \\ -2 \\ 1 \end{bmatrix}$$

Thse vectors can't possibly be linearly independent. Tell why. Next obtain a linearly independent subset of these vectors which has the same span as these vectors. In other words, find a basis for the span of these vectors.

Exercise 4.10.26 *Here are some vectors in* $\mathbb{R}^4$*.*

$$\begin{bmatrix} 1 \\ 3 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -\frac{3}{2} \\ -\frac{9}{2} \\ \frac{3}{2} \\ -\frac{3}{2} \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ -2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 0 \\ 1 \end{bmatrix}$$

Thse vectors can't possibly be linearly independent. Tell why. Next obtain a linearly independent subset of these vectors which has the same span as these vectors. In other words, find a basis for the span of these vectors.

Exercise 4.10.27 *Here are some vectors in* $\mathbb{R}^4$*.*

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

Thse vectors can't possibly be linearly independent. Tell why. Next obtain a linearly independent subset of these vectors which has the same span as these vectors. In other words, find a basis for the span of these vectors.

Exercise 4.10.28 *Here are some vectors in* $\mathbb{R}^4$*.*

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

Thse vectors can't possibly be linearly independent. Tell why. Next obtain a linearly independent subset of these vectors which has the same span as these vectors. In other words, find a basis for the span of these vectors.

Exercise 4.10.29 *Here are some vectors in* $\mathbb{R}^4$.

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

Thse vectors can't possibly be linearly independent. Tell why. Next obtain a linearly independent subset of these vectors which has the same span as these vectors. In other words, find a basis for the span of these vectors.

Exercise 4.10.30 *Here are some vectors in* $\mathbb{R}^4$.

$$\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 3 \\ 3 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -9 \\ -2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Thse vectors can't possibly be linearly independent. Tell why. Next obtain a linearly independent subset of these vectors which has the same span as these vectors. In other words, find a basis for the span of these vectors.

Exercise 4.10.31 *Here are some vectors in* $\mathbb{R}^4$.

$$\begin{bmatrix} 1 \\ b+1 \\ a \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 3b+3 \\ 3a \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ b+2 \\ 2a+1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2b-5 \\ -5a-7 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ b+2 \\ 2a+2 \\ 1 \end{bmatrix}$$

Thse vectors can't possibly be linearly independent. Tell why. Next obtain a linearly independent subset of these vectors which has the same span as these vectors. In other words, find a basis for the span of these vectors.

Exercise 4.10.32 *Let* $H = span \left\{ \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \\ 3 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -2 \\ -2 \end{bmatrix} \right\}$. *Find the dimension of* H *and determine a basis.*

Exercise 4.10.33 *Let* H *denote span* $\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ -2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 5 \\ -5 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \\ -2 \end{bmatrix} \right\}$. *Find the dimension of* H *and determine a basis.*

Exercise 4.10.34 *Let* H *denote span* $\left\{ \begin{bmatrix} -2 \\ 1 \\ 1 \\ -3 \end{bmatrix}, \begin{bmatrix} -9 \\ 4 \\ 3 \\ -9 \end{bmatrix}, \begin{bmatrix} -33 \\ 15 \\ 12 \\ -36 \end{bmatrix}, \begin{bmatrix} -22 \\ 10 \\ 8 \\ -24 \end{bmatrix} \right\}$. *Find the dimension of* H *and determine a basis.*

Exercise 4.10.35 *Let H denote span* $\left\{ \begin{bmatrix} -1 \\ 1 \\ -1 \\ -2 \end{bmatrix}, \begin{bmatrix} -4 \\ 3 \\ -2 \\ -4 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \\ -1 \\ -2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -2 \\ -4 \end{bmatrix}, \begin{bmatrix} -7 \\ 5 \\ -3 \\ -6 \end{bmatrix} \right\}$. *Find the dimension of H and determine a basis.*

Exercise 4.10.36 *Let H denote span* $\left\{ \begin{bmatrix} 2 \\ 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 8 \\ 15 \\ 6 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 6 \\ 6 \\ 3 \end{bmatrix}, \begin{bmatrix} 8 \\ 15 \\ 6 \\ 3 \end{bmatrix} \right\}$. *Find the dimension of H and determine a basis.*

Exercise 4.10.37 *Let H denote span* $\left\{ \begin{bmatrix} 0 \\ 2 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 6 \\ 0 \\ -2 \end{bmatrix}, \begin{bmatrix} -2 \\ 16 \\ 0 \\ -6 \end{bmatrix}, \begin{bmatrix} -3 \\ 22 \\ 0 \\ -8 \end{bmatrix} \right\}$. *Find the dimension of H and determine a basis.*

Exercise 4.10.38 *Let H denote span* $\left\{ \begin{bmatrix} 5 \\ 1 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 14 \\ 3 \\ 2 \\ 8 \end{bmatrix}, \begin{bmatrix} 38 \\ 8 \\ 6 \\ 24 \end{bmatrix}, \begin{bmatrix} 47 \\ 10 \\ 7 \\ 28 \end{bmatrix}, \begin{bmatrix} 10 \\ 2 \\ 3 \\ 12 \end{bmatrix} \right\}$. *Find the dimension of H and determine a basis.*

Exercise 4.10.39 *Let H denote span* $\left\{ \begin{bmatrix} 6 \\ 1 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 17 \\ 3 \\ 2 \\ 10 \end{bmatrix}, \begin{bmatrix} 52 \\ 9 \\ 7 \\ 35 \end{bmatrix}, \begin{bmatrix} 18 \\ 3 \\ 4 \\ 20 \end{bmatrix} \right\}$. *Find the dimension of H and determine a basis.*

Exercise 4.10.40 *Let* $M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \in \mathbb{R}^4 : \sin(u_1) = 1 \right\}$. *Is M a subspace? Explain.*

Exercise 4.10.41 *Let* $M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \in \mathbb{R}^4 : \|u_1\| \leq 4 \right\}$. *Is M a subspace? Explain.*

Exercise 4.10.42 *Let* $M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \in \mathbb{R}^4 : u_i \geq 0 \text{ for each } i = 1,2,3,4 \right\}$. *Is M a subspace? Explain.*

Exercise 4.10.43 *Let $\vec{w}, \vec{w}_1$ be given vectors in $\mathbb{R}^4$ and define*

$$M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \in \mathbb{R}^4 : \vec{w} \bullet \vec{u} = 0 \text{ and } \vec{w}_1 \bullet \vec{u} = 0 \right\}.$$

Is M a subspace? Explain.

Exercise 4.10.44 *Let $\vec{w} \in \mathbb{R}^4$ and let $M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \in \mathbb{R}^4 : \vec{w} \bullet \vec{u} = 0 \right\}$. Is M a subspace? Explain.*

Exercise 4.10.45 *Let $M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \in \mathbb{R}^4 : u_3 \geq u_1 \right\}$. Is M a subspace? Explain.*

Exercise 4.10.46 *Let $M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \in \mathbb{R}^4 : u_3 = u_1 = 0 \right\}$. Is M a subspace? Explain.*

Exercise 4.10.47 *Consider the set of vectors S given by*

$$S = \left\{ \begin{bmatrix} 4u + v - 5w \\ 12u + 6v - 6w \\ 4u + 4v + 4w \end{bmatrix} : u, v, w \in \mathbb{R} \right\}.$$

Is S a subspace of $\mathbb{R}^3$? If so, explain why, give a basis for the subspace and find its dimension.

Exercise 4.10.48 *Consider the set of vectors S given by*

$$S = \left\{ \begin{bmatrix} 2u + 6v + 7w \\ -3u - 9v - 12w \\ 2u + 6v + 6w \\ u + 3v + 3w \end{bmatrix} : u, v, w \in \mathbb{R} \right\}.$$

Is S a subspace of $\mathbb{R}^4$? If so, explain why, give a basis for the subspace and find its dimension.

Exercise 4.10.49 *Consider the set of vectors S given by*

$$S = \left\{ \begin{bmatrix} 2u + v \\ 6v - 3u + 3w \\ 3v - 6u + 3w \end{bmatrix} : u, v, w \in \mathbb{R} \right\}.$$

Is this set of vectors a subspace of $\mathbb{R}^3$? If so, explain why, give a basis for the subspace and find its dimension.

Exercise 4.10.50 *Consider the vectors of the form*

$$\left\{ \left[\begin{array}{c} 2u+v+7w \\ u-2v+w \\ -6v-6w \end{array} \right] : u,v,w \in \mathbb{R} \right\}.$$

Is this set of vectors a subspace of $\mathbb{R}^3$*? If so, explain why, give a basis for the subspace and find its dimension.*

Exercise 4.10.51 *Consider the vectors of the form*

$$\left\{ \left[\begin{array}{c} 3u+v+11w \\ 18u+6v+66w \\ 28u+8v+100w \end{array} \right] : u,v,w \in \mathbb{R} \right\}.$$

Is this set of vectors a subspace of $\mathbb{R}^3$*? If so, explain why, give a basis for the subspace and find its dimension.*

Exercise 4.10.52 *Consider the vectors of the form*

$$\left\{ \left[\begin{array}{c} 3u+v \\ 2w-4u \\ 2w-2v-8u \end{array} \right] : u,v,w \in \mathbb{R} \right\}.$$

Is this set of vectors a subspace of $\mathbb{R}^3$*? If so, explain why, give a basis for the subspace and find its dimension.*

Exercise 4.10.53 *Consider the set of vectors S given by*

$$\left\{ \left[\begin{array}{c} u+v+w \\ 2u+2v+4w \\ u+v+w \\ 0 \end{array} \right] : u,v,w \in \mathbb{R} \right\}.$$

Is S a subspace of $\mathbb{R}^4$*? If so, explain why, give a basis for the subspace and find its dimension.*

Exercise 4.10.54 *Consider the set of vectors S given by*

$$\left\{ \left[\begin{array}{c} v \\ -3u-3w \\ 8u-4v+4w \end{array} \right] : u,v,w \in \mathbb{R} \right\}.$$

Is S a subspace of $\mathbb{R}^3$*? If so, explain why, give a basis for the subspace and find its dimension.*

Exercise 4.10.55 *If you have* 5 *vectors in* $\mathbb{R}^5$ *and the vectors are linearly independent, can it always be concluded they span* $\mathbb{R}^5$*? Explain.*

Exercise 4.10.56 *If you have* 6 *vectors in* $\mathbb{R}^5$*, is it possible they are linearly independent? Explain.*

Exercise 4.10.57 *Suppose A is an $m \times n$ matrix and $\{\vec{w}_1, \cdots, \vec{w}_k\}$ is a linearly independent set of vectors in $A(\mathbb{R}^n) \subseteq \mathbb{R}^m$. Now suppose $A\vec{z}_i = \vec{w}_i$. Show $\{\vec{z}_1, \cdots, \vec{z}_k\}$ is also independent.*

Exercise 4.10.58 *Suppose V, W are subspaces of $\mathbb{R}^n$. Let $V \cap W$ be all vectors which are in both V and W. Show that $V \cap W$ is a subspace also.*

Exercise 4.10.59 *Suppose V and W both have dimension equal to 7 and they are subspaces of $\mathbb{R}^{10}$. What are the possibilities for the dimension of $V \cap W$? **Hint:** Remember that a linear independent set can be extended to form a basis.*

Exercise 4.10.60 *Suppose V has dimension p and W has dimension q and they are each contained in a subspace, U which has dimension equal to n where $n > \max(p, q)$. What are the possibilities for the dimension of $V \cap W$? **Hint:** Remember that a linearly independent set can be extended to form a basis.*

Exercise 4.10.61 *Suppose A is an $m \times n$ matrix and B is an $n \times p$ matrix. Show that*

$$\dim(\ker(AB)) \leq \dim(\ker(A)) + \dim(\ker(B)).$$

***Hint:** Consider the subspace, $B(\mathbb{R}^p) \cap \ker(A)$ and suppose a basis for this subspace is $\{\vec{w}_1, \cdots, \vec{w}_k\}$. Now suppose $\{\vec{u}_1, \cdots, \vec{u}_r\}$ is a basis for $\ker(B)$. Let $\{\vec{z}_1, \cdots, \vec{z}_k\}$ be such that $B\vec{z}_i = \vec{w}_i$ and argue that*

$$\ker(AB) \subseteq \textit{span}\{\vec{u}_1, \cdots, \vec{u}_r, \vec{z}_1, \cdots, \vec{z}_k\}.$$

Exercise 4.10.62 *Show that if A is an $m \times n$ matrix, then $\ker(A)$ is a subspace of $\mathbb{R}^n$.*

Exercise 4.10.63 *Find the rank of the following matrix. Also find a basis for the row and column spaces.*

$$\begin{bmatrix} 1 & 3 & 0 & -2 & 0 & 3 \\ 3 & 9 & 1 & -7 & 0 & 8 \\ 1 & 3 & 1 & -3 & 1 & -1 \\ 1 & 3 & -1 & -1 & -2 & 10 \end{bmatrix}$$

Exercise 4.10.64 *Find the rank of the following matrix. Also find a basis for the row and column spaces.*

$$\begin{bmatrix} 1 & 3 & 0 & -2 & 7 & 3 \\ 3 & 9 & 1 & -7 & 23 & 8 \\ 1 & 3 & 1 & -3 & 9 & 2 \\ 1 & 3 & -1 & -1 & 5 & 4 \end{bmatrix}$$

Exercise 4.10.65 *Find the rank of the following matrix. Also find a basis for the row and column spaces.*

$$\begin{bmatrix} 1 & 0 & 3 & 0 & 7 & 0 \\ 3 & 1 & 10 & 0 & 23 & 0 \\ 1 & 1 & 4 & 1 & 7 & 0 \\ 1 & -1 & 2 & -2 & 9 & 1 \end{bmatrix}$$

Exercise 4.10.66 *Find the rank of the following matrix. Also find a basis for the row and column spaces.*

$$\begin{bmatrix} 1 & 0 & 3 \\ 3 & 1 & 10 \\ 1 & 1 & 4 \\ 1 & -1 & 2 \end{bmatrix}$$

Exercise 4.10.67 *Find the rank of the following matrix. Also find a basis for the row and column spaces.*

$$\begin{bmatrix} 0 & 0 & -1 & 0 & 1 \\ 1 & 2 & 3 & -2 & -18 \\ 1 & 2 & 2 & -1 & -11 \\ -1 & -2 & -2 & 1 & 11 \end{bmatrix}$$

Exercise 4.10.68 *Find the rank of the following matrix. Also find a basis for the row and column spaces.*

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

Exercise 4.10.69 *Find* $\ker(A)$ *for the following matrices.*

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

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

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

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

4.11 Orthogonality and the Gram Schmidt Process

Outcomes

A. *Determine if a given set is orthogonal or orthonormal.*

B. *Determine if a given matrix is orthogonal.*

C. *Given a linearly independent set, use the Gram-Schmidt Process to find corresponding orthogonal and orthonormal sets.*

D. *Find the orthogonal projection of a vector onto a subspace.*

E. *Find the least squares approximation for a collection of points.*

4.11.1. Orthogonal and Orthonormal Sets

In this section, we examine what it means for vectors (and sets of vectors) to be orthogonal and orthonormal. First, it is necessary to review some important concepts. You may recall the definitions for the span of a set of vectors and a linear independent set of vectors. We include the definitions and examples here for convenience.

Definition 4.114: Span of a Set of Vectors and Subspace

The collection of all linear combinations of a set of vectors $\{\vec{u}_1, \cdots, \vec{u}_k\}$ *in* $\mathbb{R}^n$ *is known as the span of these vectors and is written as* span$\{\vec{u}_1, \cdots, \vec{u}_k\}$.
We call a collection of the form span$\{\vec{u}_1, \cdots, \vec{u}_k\}$ *a subspace of* $\mathbb{R}^n$.

Consider the following example.

Example 4.115: Span of Vectors

Describe the span of the vectors $\vec{u} = \left[\begin{array}{ccc} 1 & 1 & 0 \end{array}\right]^T$ *and* $\vec{v} = \left[\begin{array}{ccc} 3 & 2 & 0 \end{array}\right]^T \in \mathbb{R}^3$.

Solution. You can see that any linear combination of the vectors $\vec{u}$ and $\vec{v}$ yields a vector $\left[\begin{array}{ccc} x & y & 0 \end{array}\right]^T$ in the XY-plane.

Moreover every vector in the XY-plane is in fact such a linear combination of the vectors

$\vec{u}$ and $\vec{v}$. That's because

$$\begin{bmatrix} x \\ y \\ 0 \end{bmatrix} = (-2x+3y)\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + (x-y)\begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}$$

Thus $\operatorname{span}\{\vec{u}, \vec{v}\}$ is precisely the XY-plane. ☐

The span of a set of a vectors in $\mathbb{R}^n$ is what we call a **subspace of** $\mathbb{R}^n$. A subspace W is characterized by the feature that any linear combination of vectors of W is again a vector contained in W.

Another important property of sets of vectors is called linear independence.

Definition 4.116: Linearly Independent Set of Vectors

*A set of non-zero vectors $\{\vec{u}_1, \cdots, \vec{u}_k\}$ in $\mathbb{R}^n$ is said to be **linearly independent** if no vector in that set is in the span of the other vectors of that set.*

Here is an example.

Example 4.117: Linearly Independent Vectors

Consider vectors $\vec{u} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T$, $\vec{v} = \begin{bmatrix} 3 & 2 & 0 \end{bmatrix}^T$, and $\vec{w} = \begin{bmatrix} 4 & 5 & 0 \end{bmatrix}^T \in \mathbb{R}^3$. Verify whether the set $\{\vec{u}, \vec{v}, \vec{w}\}$ is linearly independent.

Solution. We already verified in Example 4.115 that $\operatorname{span}\{\vec{u}, \vec{v}\}$ is the XY-plane. Since $\vec{w}$ is clearly also in the XY-plane, then the set $\{\vec{u}, \vec{v}, \vec{w}\}$ is **not** linearly independent. ☐

In terms of spanning, a set of vectors is linearly independent if it does not contain unnecessary vectors. In the previous example you can see that the vector $\vec{w}$ does not help to span any new vector not already in the span of the other two vectors. However you can verify that the set $\{\vec{u}, \vec{v}\}$ is linearly independent, since you will not get the XY-plane as the span of a single vector.

We can also determine if a set of vectors is linearly independent by examining linear combinations. A set of vectors is linearly independent if and only if whenever a linear combination of these vectors equals zero, it follows that all the coefficients equal zero. It is a good exercise to verify this equivalence, and this latter condition is often used as the (equivalent) definition of linear independence.

If a subspace is spanned by a linearly independent set of vectors, then we say that it is a basis for the subspace.

Definition 4.118: Basis of a Subspace

Let V be a subspace of $\mathbb{R}^n$. Then $\{\vec{u}_1, \cdots, \vec{u}_k\}$ is a ***basis*** *for V if the following two conditions hold.*

1. *span* $\{\vec{u}_1, \cdots, \vec{u}_k\} = V$
2. $\{\vec{u}_1, \cdots, \vec{u}_k\}$ *is linearly independent*

Thus the set of vectors $\{\vec{u}, \vec{v}\}$ from Example 4.117 is a basis for XY-plane in $\mathbb{R}^3$ since it is both linearly independent and spans the XY-plane.

Recall from the properties of the dot product of vectors that two vectors $\vec{u}$ and $\vec{v}$ are orthogonal if $\vec{u} \bullet \vec{v} = 0$. Suppose a vector is orthogonal to a spanning set of $\mathbb{R}^n$. What can be said about such a vector? This is the discussion in the following example.

Example 4.119: Orthogonal Vector to a Spanning Set

Let $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\} \in \mathbb{R}^n$ and suppose $\mathbb{R}^n = \text{span}\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$. Furthermore, suppose that there exists a vector $\vec{u} \in \mathbb{R}^n$ for which $\vec{u} \bullet \vec{x}_j = 0$ for all j, $1 \leq j \leq k$. What type of vector is $\vec{u}$?

Solution. Write $\vec{u} = t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k$ for some $t_1, t_2, \ldots, t_k \in \mathbb{R}$ (this is possible because $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$ span $\mathbb{R}^n$).

Then

$$\begin{aligned}
\|\vec{u}\|^2 &= \vec{u} \bullet \vec{u} \\
&= \vec{u} \bullet (t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k) \\
&= \vec{u} \bullet (t_1\vec{x}_1) + \vec{u} \bullet (t_2\vec{x}_2) + \cdots + \vec{u} \bullet (t_k\vec{x}_k) \\
&= t_1(\vec{u} \bullet \vec{x}_1) + t_2(\vec{u} \bullet \vec{x}_2) + \cdots + t_k(\vec{u} \bullet \vec{x}_k) \\
&= t_1(0) + t_2(0) + \cdots + t_k(0) = 0.
\end{aligned}$$

Since $\|\vec{u}\|^2 = 0$, $\|\vec{u}\| = 0$. We know that $\|\vec{u}\| = 0$ if and only if $\vec{u} = \vec{0}_n$. Therefore, $\vec{u} = \vec{0}_n$. In conclusion, the only vector orthogonal to every vector of a spanning set of $\mathbb{R}^n$ is the zero vector. ◻

We can now discuss what is meant by an orthogonal set of vectors.

Definition 4.120: Orthogonal Set of Vectors

Let $\{\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_m\}$ be a set of vectors in $\mathbb{R}^n$. Then this set is called an ***orthogonal set*** *if the following conditions hold:*

1. $\vec{u}_i \bullet \vec{u}_j = 0$ *for all* $i \neq j$
2. $\vec{u}_i \neq \vec{0}$ *for all* i

If we have an orthogonal set of vectors and normalize each vector so they have length 1, the resulting set is called an **orthonormal set** of vectors. They can be described as follows.

Definition 4.121: Orthonormal Set of Vectors

A set of vectors, $\{\vec{w}_1, \cdots, \vec{w}_m\}$ is said to be an ***orthonormal*** *set if*

$$\vec{w}_i \bullet \vec{w}_j = \delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

Note that all orthonormal sets are orthogonal, but the reverse is not necessarily true since the vectors may not be normalized. In order to normalize the vectors, we simply need divide each one by its length.

Definition 4.122: Normalizing an Orthogonal Set

Normalizing an orthogonal set is the process of turning an orthogonal (but not orthonormal) set into an orthonormal set. If $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is an orthogonal subset of $\mathbb{R}^n$, then

$$\left\{ \frac{1}{\|\vec{u}_1\|}\vec{u}_1, \frac{1}{\|\vec{u}_2\|}\vec{u}_2, \ldots, \frac{1}{\|\vec{u}_k\|}\vec{u}_k \right\}$$

is an orthonormal set.

We illustrate this concept in the following example.

Example 4.123: Orthonormal Set

Consider the set of vectors given by

$$\{\vec{u}_1, \vec{u}_2\} = \left\{ \left[\begin{array}{r} 1 \\ 1 \end{array} \right], \left[\begin{array}{r} -1 \\ 1 \end{array} \right] \right\}$$

Show that it is an orthogonal set of vectors but not an orthonormal one. Find the corresponding orthonormal set.

Solution. One easily verifies that $\vec{u}_1 \bullet \vec{u}_2 = 0$ and $\{\vec{u}_1, \vec{u}_2\}$ is an orthogonal set of vectors. On the other hand one can compute that $\|\vec{u}_1\| = \|\vec{u}_2\| = \sqrt{2} \neq 1$ and thus it is not an orthonormal set.

Thus to find a corresponding orthonormal set, we simply need to normalize each vector.

We will write $\{\vec{w}_1, \vec{w}_2\}$ for the corresponding orthonormal set. Then,

$$\begin{aligned} \vec{w}_1 &= \frac{1}{\|\vec{u}_1\|}\vec{u}_1 \\ &= \frac{1}{\sqrt{2}}\left[\begin{array}{r} 1 \\ 1 \end{array}\right] \\ &= \left[\begin{array}{r} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{array}\right] \end{aligned}$$

Similarly,

$$\begin{aligned} \vec{w}_2 &= \frac{1}{\|\vec{u}_2\|}\vec{u}_2 \\ &= \frac{1}{\sqrt{2}}\left[\begin{array}{r} -1 \\ 1 \end{array}\right] \\ &= \left[\begin{array}{r} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{array}\right] \end{aligned}$$

Therefore the corresponding orthonormal set is

$$\{\vec{w}_1, \vec{w}_2\} = \left\{ \left[\begin{array}{r} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{array}\right], \left[\begin{array}{r} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{array}\right] \right\}$$

You can verify that this set is orthogonal. ◻

Consider an orthogonal set of vectors in $\mathbb{R}^n$, written $\{\vec{w}_1, \cdots, \vec{w}_k\}$ with $k \leq n$. The span of these vectors is a subspace W of $\mathbb{R}^n$. If we could show that this orthogonal set is also linearly independent, we would have a basis of W. We will show this in the next theorem.

Theorem 4.124: Orthogonal Basis of a Subspace

Let $\{\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_k\}$ be an orthonormal set of vectors in $\mathbb{R}^n$. Then this set is linearly independent and forms a basis for the subspace $W = span\{\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_k\}$.

Proof. To show it is a linearly independent set, suppose a linear combination of these vectors equals $\vec{0}$, such as:

$$a_1\vec{w}_1 + a_2\vec{w}_2 + \cdots + a_k\vec{w}_k = \vec{0}, a_i \in \mathbb{R}$$

We need to show that all $a_i = 0$. To do so, take the dot product of each side of the above equation with the vector $\vec{w}_i$ and obtain the following.

$$\begin{aligned} \vec{w}_i \bullet (a_1\vec{w}_1 + a_2\vec{w}_2 + \cdots + a_k\vec{w}_k) &= \vec{w}_i \bullet \vec{0} \\ a_1(\vec{w}_i \bullet \vec{w}_1) + a_2(\vec{w}_i \bullet \vec{w}_2) + \cdots + a_k(\vec{w}_i \bullet \vec{w}_k) &= 0 \end{aligned}$$

Now since the set is orthogonal, $\vec{w}_i \bullet \vec{w}_m = 0$ for all $m \neq i$, so we have:

$$a_1(0) + \cdots + a_i(\vec{w}_i \bullet \vec{w}_i) + \cdots + a_k(0) = 0$$

$$a_i \|\vec{w}_i\|^2 = 0$$

Since the set is orthogonal, we know that $\|\vec{w}_i\|^2 \neq 0$. It follows that $a_i = 0$. Since the a_i was chosen arbitrarily, the set $\{\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_k\}$ is linearly independent.

Finally since $W = \text{span}\{\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_k\}$, the set of vectors also spans W and therefore forms a basis of W.

□

If an orthogonal set is a basis for a subspace, we call this an orthogonal basis. Similarly, if an orthonormal set is a basis, we call this an orthonormal basis.

We conclude this section with a discussion of Fourier expansions. Given any orthogonal basis B of $\mathbb{R}^n$ and an arbitrary vector $\vec{x} \in \mathbb{R}^n$, how do we express $\vec{x}$ as a linear combination of vectors in B? The solution is Fourier expansion.

Theorem 4.125: Fourier Expansion

Let V be a subspace of $\mathbb{R}^n$ and suppose $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}$ is an orthogonal basis of V. Then for any $\vec{x} \in V$,

$$\vec{x} = \left(\frac{\vec{x} \bullet \vec{u}_1}{\|\vec{u}_1\|^2} \right) \vec{u}_1 + \left(\frac{\vec{x} \bullet \vec{u}_2}{\|\vec{u}_2\|^2} \right) \vec{u}_2 + \cdots + \left(\frac{\vec{x} \bullet \vec{u}_m}{\|\vec{u}_m\|^2} \right) \vec{u}_m.$$

This expression is called the Fourier expansion of $\vec{x}$, and

$$\frac{\vec{x} \bullet \vec{u}_j}{\|\vec{u}_j\|^2},$$

$j = 1, 2, \ldots, m$ are the Fourier coefficients.

Consider the following example.

Example 4.126: Fourier Expansion

Let $\vec{u}_1 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$, *and* $\vec{u}_3 = \begin{bmatrix} 5 \\ 1 \\ -2 \end{bmatrix}$, *and let* $\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.
Then $B = \{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ is an orthogonal basis of $\mathbb{R}^3$.
Compute the Fourier expansion of $\vec{x}$, thus writing $\vec{x}$ as a linear combination of the vectors of B.

Solution. Since B is a basis (verify!) there is a unique way to express $\vec{x}$ as a linear combination of the vectors of B. Moreover since B is an orthogonal basis (verify!), then this can be done by computing the Fourier expansion of $\vec{x}$.

That is:

$$\vec{x} = \left(\frac{\vec{x} \bullet \vec{u}_1}{\|\vec{u}_1\|^2}\right)\vec{u}_1 + \left(\frac{\vec{x} \bullet \vec{u}_2}{\|\vec{u}_2\|^2}\right)\vec{u}_2 + \left(\frac{\vec{x} \bullet \vec{u}_3}{\|\vec{u}_3\|^2}\right)\vec{u}_3.$$

We readily compute:

$$\frac{\vec{x} \bullet \vec{u}_1}{\|\vec{u}_1\|^2} = \frac{2}{6}, \ \frac{\vec{x} \bullet \vec{u}_2}{\|\vec{u}_2\|^2} = \frac{3}{5}, \text{ and } \frac{\vec{x} \bullet \vec{u}_3}{\|\vec{u}_3\|^2} = \frac{4}{30}.$$

Therefore,

$$\left[\begin{array}{r} 1 \\ 1 \\ 1 \end{array}\right] = \frac{1}{3}\left[\begin{array}{r} 1 \\ -1 \\ 2 \end{array}\right] + \frac{3}{5}\left[\begin{array}{r} 0 \\ 2 \\ 1 \end{array}\right] + \frac{2}{15}\left[\begin{array}{r} 5 \\ 1 \\ -2 \end{array}\right].$$

□

4.11.2. Orthogonal Matrices

Recall that the process to find the inverse of a matrix was often cumbersome. In contrast, it was very easy to take the transpose of a matrix. Luckily for some special matrices, the transpose equals the inverse. When an $n \times n$ matrix has all real entries and its transpose equals its inverse, the matrix is called an **orthogonal matrix**.

The precise definition is as follows.

Definition 4.127: Orthogonal Matrices

*A real $n \times n$ matrix U is called an **orthogonal** matrix if $UU^T = U^TU = I$.*

Note since U is assumed to be a square matrix, it suffices to verify only one of these equalities $UU^T = I$ or $U^TU = I$ holds to guarantee that U^T is the inverse of U.

Consider the following example.

Example 4.128: Orthogonal Matrix

Show the matrix

$$U = \left[\begin{array}{rr} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{array}\right]$$

is orthogonal.

Solution. All we need to do is verify (one of the equations from) the requirements of Definition 4.127.

$$UU^T = \left[\begin{array}{rr} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{array}\right]\left[\begin{array}{rr} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{array}\right] = \left[\begin{array}{rr} 1 & 0 \\ 0 & 1 \end{array}\right]$$

Since $UU^T = I$, this matrix is orthogonal. ☐

Here is another example.

Example 4.129: Orthogonal Matrix

Let $U = \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{array} \right]$. *Is U orthogonal?*

Solution. Again the answer is yes and this can be verified simply by showing that $U^T U = I$:

$$\begin{aligned} U^T U &= \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{array} \right]^T \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{array} \right] \\ &= \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{array} \right] \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{array} \right] \\ &= \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right] \end{aligned}$$

☐

When we say that U is orthogonal, we are saying that $UU^T = I$, meaning that

$$\sum_j u_{ij} u_{jk}^T = \sum_j u_{ij} u_{kj} = \delta_{ik}$$

where δ_{ij} is the **Kronecker symbol** defined by

$$\delta_{ij} = \left\{ \begin{array}{l} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{array} \right.$$

In words, the product of the i^{th} row of U with the k^{th} row gives 1 if $i = k$ and 0 if $i \neq k$. The same is true of the columns because $U^T U = I$ also. Therefore,

$$\sum_j u_{ij}^T u_{jk} = \sum_j u_{ji} u_{jk} = \delta_{ik}$$

which says that the product of one column with another column gives 1 if the two columns are the same and 0 if the two columns are different.

More succinctly, this states that if $\vec{u}_1, \cdots, \vec{u}_n$ are the columns of U, an orthogonal matrix, then

$$\vec{u}_i \bullet \vec{u}_j = \delta_{ij} = \left\{ \begin{array}{l} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{array} \right.$$

We will say that the columns form an orthonormal set of vectors, and similarly for the rows. Thus a matrix is **orthogonal** if its rows (or columns) form an **orthonormal** set of vectors. Notice that the convention is to call such a matrix orthogonal rather than orthonormal (although this may make more sense!).

Proposition 4.130: Orthonormal Basis

The rows of an $n \times n$ orthogonal matrix form an orthonormal basis of $\mathbb{R}^n$. Further, any orthonormal basis of $\mathbb{R}^n$ can be used to construct an $n \times n$ orthogonal matrix.

Proof. Recall from Theorem 4.124 that an orthonormal set is linearly independent and forms a basis for its span. Since the rows of an $n \times n$ orthogonal matrix form an orthonormal set, they must be linearly independent. Now we have n linearly independent vectors, and it follows that their span equals $\mathbb{R}^n$. Therefore these vectors form an orthonormal basis for $\mathbb{R}^n$.

Suppose now that we have an orthonormal basis for $\mathbb{R}^n$. Since the basis will contain n vectors, these can be used to construct an $n \times n$ matrix, with each vector becoming a row. Therefore the matrix is composed of orthonormal rows, which by our above discussion, means that the matrix is orthogonal. Note we could also have construct a matrix with each vector becoming a column instead, and this would again be an orthogonal matrix. In fact this is simply the transpose of the previous matrix. ☐

Consider the following proposition.

Proposition 4.131: Determinant of Orthogonal Matrices

Suppose U is an orthogonal matrix. Then $\det(U) = \pm 1$.

Proof. This result follows from the properties of determinants. Recall that for any matrix A, $\det(A)^T = \det(A)$. Now if U is orthogonal, then:

$$(\det(U))^2 = \det\left(U^T\right)\det(U) = \det\left(U^T U\right) = \det(I) = 1$$

Therefore $(\det(U))^2 = 1$ and it follows that $\det(U) = \pm 1$. ☐

Orthogonal matrices are divided into two classes, proper and improper. The proper orthogonal matrices are those whose determinant equals 1 and the improper ones are those whose determinant equals -1. The reason for the distinction is that the improper orthogonal matrices are sometimes considered to have no physical significance. These matrices cause a change in orientation which would correspond to material passing through itself in a non physical manner. Thus in considering which coordinate systems must be considered in certain applications, you only need to consider those which are related by a proper orthogonal transformation. Geometrically, the linear transformations determined by the proper orthogonal matrices correspond to the composition of rotations.

We conclude this section with two useful properties of orthogonal matrices.

Example 4.132: Product and Inverse of Orthogonal Matrices

Suppose A and B are orthogonal matrices. Then AB and A^{-1} both exist and are orthogonal.

Solution. First we examine the product AB.

$$(AB)(B^T A^T) = A(BB^T)A^T = AA^T = I$$

Since AB is square, $B^T A^T = (AB)^T$ is the inverse of AB, so AB is invertible, and $(AB)^{-1} = (AB)^T$ Therefore, AB is orthogonal.

Next we show that $A^{-1} = A^T$ is also orthogonal.

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

Therefore A^{-1} is also orthogonal. □

4.11.3. Gram-Schmidt Process

The Gram-Schmidt process is an algorithm to transform a set of vectors into an orthonormal set spanning the same subspace, that is generating the same collection of linear combinations (see Definition 4.7).

The goal of the Gram-Schmidt process is to take a linearly independent set of vectors and transform it into an orthonormal set with the same span. The first objective is to construct an orthogonal set of vectors with the same span, since from there an orthonormal set can be obtained by simply dividing each vector by its length.

Algorithm 4.133: Gram-Schmidt Process

Let $\{\vec{u}_1, \cdots, \vec{u}_n\}$ be a set of linearly independent vectors in $\mathbb{R}^n$.
I: *Construct a new set of vectors $\{\vec{v}_1, \cdots, \vec{v}_n\}$ as follows:*

$$\begin{aligned}
\vec{v}_1 &= \vec{u}_1 \\
\vec{v}_2 &= \vec{u}_2 - \left(\frac{\vec{u}_2 \bullet \vec{v}_1}{\|\vec{v}_1\|^2}\right)\vec{v}_1 \\
\vec{v}_3 &= \vec{u}_3 - \left(\frac{\vec{u}_3 \bullet \vec{v}_1}{\|\vec{v}_1\|^2}\right)\vec{v}_1 - \left(\frac{\vec{u}_3 \bullet \vec{v}_2}{\|\vec{v}_2\|^2}\right)\vec{v}_2 \\
&\vdots \\
\vec{v}_n &= \vec{u}_n - \left(\frac{\vec{u}_n \bullet \vec{v}_1}{\|\vec{v}_1\|^2}\right)\vec{v}_1 - \left(\frac{\vec{u}_n \bullet \vec{v}_2}{\|\vec{v}_2\|^2}\right)\vec{v}_2 - \cdots - \left(\frac{\vec{u}_n \bullet \vec{v}_{n-1}}{\|\vec{v}_{n-1}\|^2}\right)\vec{v}_{n-1}
\end{aligned}$$

II: *Now let $\vec{w}_i = \dfrac{\vec{v}_i}{\|\vec{v}_i\|}$ for $i = 1, \cdots, n$.*
Then

1. *$\{\vec{v}_1, \cdots, \vec{v}_n\}$ is an orthogonal set.*
2. *$\{\vec{w}_1, \cdots, \vec{w}_n\}$ is an orthonormal set.*
3. *span $\{\vec{u}_1, \cdots, \vec{u}_n\}$ = span $\{\vec{v}_1, \cdots, \vec{v}_n\}$ = span $\{\vec{w}_1, \cdots, \vec{w}_n\}$.*

Proof. The full proof of this algorithm is beyond this material, however here is an indication of the arguments.

To show that $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is an orthogonal set, let

$$a_2 = \frac{\vec{u}_2 \bullet \vec{v}_1}{\|\vec{v}_1\|^2}$$

then:

$$\begin{aligned}
\vec{v}_1 \bullet \vec{v}_2 &= \vec{v}_1 \bullet (\vec{u}_2 - a_2\vec{v}_1) \\
&= \vec{v}_1 \bullet \vec{u}_2 - a_2(\vec{v}_1 \bullet \vec{v}_1 \\
&= \vec{v}_1 \bullet \vec{u}_2 - \frac{\vec{u}_2 \bullet \vec{v}_1}{\|\vec{v}_1\|^2}\|\vec{v}_1\|^2 \\
&= (\vec{v}_1 \bullet \vec{u}_2) - (\vec{u}_2 \bullet \vec{v}_1) = 0
\end{aligned}$$

Now that you have shown that $\{\vec{v}_1, \vec{v}_2\}$ is orthogonal, use the same method as above to show that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is also orthogonal, and so on.

Then in a similar fashion you show that span $\{\vec{u}_1, \cdots, \vec{u}_n\}$ = span $\{\vec{v}_1, \cdots, \vec{v}_n\}$.

Finally defining $\vec{w}_i = \dfrac{\vec{v}_i}{\|\vec{v}_i\|}$ for $i = 1, \cdots, n$ does not affect orthogonality and yields vectors of length 1, hence an orthonormal set. You can also observe that it does not affect the span either and the proof would be complete. □

Consider the following example.

Example 4.134: Find Orthonormal Set with Same Span

Consider the set of vectors $\{\vec{u}_1, \vec{u}_2\}$ *given as in Example 4.115. That is*

$$\vec{u}_1 = \left[\begin{array}{c} 1 \\ 1 \\ 0 \end{array} \right], \vec{u}_2 = \left[\begin{array}{c} 3 \\ 2 \\ 0 \end{array} \right] \in \mathbb{R}^3$$

Use the Gram-Schmidt algorithm to find an orthonormal set of vectors $\{\vec{w}_1, \vec{w}_2\}$ *having the same span.*

Solution. We already remarked that the set of vectors in $\{\vec{u}_1, \vec{u}_2\}$ is linearly independent, so we can proceed with the Gram-Schmidt algorithm:

$$\begin{aligned} \vec{v}_1 &= \vec{u}_1 = \left[\begin{array}{c} 1 \\ 1 \\ 0 \end{array} \right] \\ \vec{v}_2 &= \vec{u}_2 - \left(\frac{\vec{u}_2 \bullet \vec{v}_1}{\|\vec{v}_1\|^2} \right) \vec{v}_1 \\ &= \left[\begin{array}{c} 3 \\ 2 \\ 0 \end{array} \right] - \frac{5}{2} \left[\begin{array}{c} 1 \\ 1 \\ 0 \end{array} \right] \\ &= \left[\begin{array}{r} \frac{1}{2} \\ -\frac{1}{2} \\ 0 \end{array} \right] \end{aligned}$$

Now to normalize simply let

$$\vec{w}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \left[\begin{array}{r} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{array} \right]$$

$$\vec{w}_2 = \frac{\vec{v}_2}{\|\vec{v}_2\|} = \left[\begin{array}{r} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{array} \right]$$

You can verify that $\{\vec{w}_1, \vec{w}_2\}$ is an orthonormal set of vectors having the same span as $\{\vec{u}_1, \vec{u}_2\}$, namely the XY-plane. ◻

In this example, we began with a linearly independent set and found an orthonormal set of vectors which had the same span. It turns out that if we start with a basis of a subspace and apply the Gram-Schmidt algorithm, the result will be an orthogonal basis of the same subspace. We examine this in the following example.

Example 4.135: Find a Corresponding Orthogonal Basis

Let

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \vec{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \text{ and } \vec{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

and let $U = span\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$. Use the Gram-Schmidt Process to construct an orthogonal basis B of U.

Solution. First $\vec{f}_1 = \vec{x}_1$.

Next,

$$\vec{f}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Finally,

$$\vec{f}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \frac{0}{1}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1 \\ -1/2 \\ 0 \end{bmatrix}.$$

Therefore,

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 1 \\ -1/2 \\ 0 \end{bmatrix} \right\}$$

is an orthogonal basis of U. However, it is sometimes more convenient to deal with vectors having integer entries, in which case we take

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix} \right\}.$$

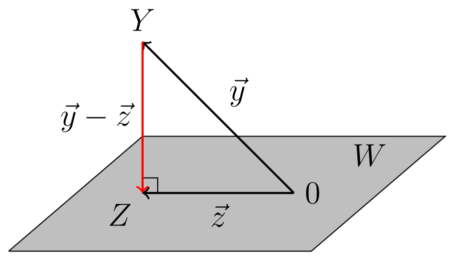
□

4.11.4. Orthogonal Projections

An important use of the Gram-Schmidt Process is in orthogonal projections, the focus of this section.

You may recall that a subspace of $\mathbb{R}^n$ is a set of vectors which contains the zero vector, and is closed under addition and scalar multiplication. Let's call such a subspace W. In particular, a plane in $\mathbb{R}^n$ which contains the origin, $(0, 0, \cdots, 0)$, is a subspace of $\mathbb{R}^n$.

Suppose a point Y in $\mathbb{R}^n$ is not contained in W, then what point Z in W is closest to Y? Using the Gram-Schmidt Process, we can find such a point. Let $\vec{y}, \vec{z}$ represent the position vectors of the points Y and Z respectively, with $\vec{y}-\vec{z}$ representing the vector connecting the two points Y and Z. It will follow that if Z is the point on W closest to Y, then $\vec{y}-\vec{z}$ will be perpendicular to W (can you see why?); in other words, $\vec{y}-\vec{z}$ is orthogonal to W (and to every vector contained in W) as in the following diagram.



The vector $\vec{z}$ is called the **orthogonal projection** of $\vec{y}$ on W. The definition is given as follows.

Definition 4.136: Orthogonal Projection

Let W be a subspace of $\mathbb{R}^n$, and Y be any point in $\mathbb{R}^n$. Then the orthogonal projection of Y onto W is given by

$$\vec{z} = proj_W(\vec{y}) = \left(\frac{\vec{y} \bullet \vec{w}_1}{\|\vec{w}_1\|^2}\right)\vec{w}_1 + \left(\frac{\vec{y} \bullet \vec{w}_2}{\|\vec{w}_2\|^2}\right)\vec{w}_2 + \cdots + \left(\frac{\vec{y} \bullet \vec{w}_m}{\|\vec{w}_m\|^2}\right)\vec{w}_m$$

where $\{\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_m\}$ is any orthogonal basis of W.

Therefore, in order to find the orthogonal projection, we must first find an orthogonal basis for the subspace. Note that one could use an orthonormal basis, but it is not necessary in this case since as you can see above the normalization of each vector is included in the formula for the projection.

Before we explore this further through an example, we show that the orthogonal projection does indeed yield a point Z (the point whose position vector is the vector $\vec{z}$ above) which is the point of W closest to Y.

Theorem 4.137: Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$ and Y any point in $\mathbb{R}^n$. Let Z be the point whose position vector is the orthogonal projection of Y onto W.
Then, Z is the point in W closest to Y.

Proof. First Z is certainly a point in W since it is in the span of a basis of W.

To show that Z is the point in W closest to Y, we wish to show that $|\vec{y}-\vec{z}_1| > |\vec{y}-\vec{z}|$ for all $\vec{z}_1 \neq \vec{z} \in W$. We begin by writing $\vec{y}-\vec{z}_1 = (\vec{y}-\vec{z}) + (\vec{z}-\vec{z}_1)$. Now, the vector $\vec{y}-\vec{z}$

is orthogonal to W, and $\vec{z}-\vec{z}_1$ is contained in W. Therefore these vectors are orthogonal to each other. By the Pythagorean Theorem, we have that

$$\|\vec{y}-\vec{z}_1\|^2 = \|\vec{y}-\vec{z}\|^2 + \|\vec{z}-\vec{z}_1\|^2 > \|\vec{y}-\vec{z}\|^2$$

This follows because $\vec{z} \neq \vec{z}_1$ so $\|\vec{z}-\vec{z}_1\|^2 > 0$.

Hence, $\|\vec{y}-\vec{z}_1\|^2 > \|\vec{y}-\vec{z}\|^2$. Taking the square root of each side, we obtain the desired result. ☐

Consider the following example.

Example 4.138: Orthogonal Projection

Let W be the plane through the origin given by the equation $x-2y+z=0$.
Find the point in W closest to the point $Y=(1,0,3)$.

Solution. We must first find an orthogonal basis for W. Notice that W is characterized by all points (a,b,c) where $c=2b-a$. In other words,

$$W = \left[\begin{array}{c} a \\ b \\ 2b-a \end{array}\right] = a\left[\begin{array}{r} 1 \\ 0 \\ -1 \end{array}\right] + b\left[\begin{array}{c} 0 \\ 1 \\ 2 \end{array}\right],\ a,b \in \mathbb{R}$$

We can thus write W as

$$\begin{aligned} W &= \operatorname{span}\{\vec{u}_1, \vec{u}_2\} \\ &= \operatorname{span}\left\{\left[\begin{array}{r} 1 \\ 0 \\ -1 \end{array}\right], \left[\begin{array}{c} 0 \\ 1 \\ 2 \end{array}\right]\right\} \end{aligned}$$

Notice that this span is a basis of W as it is linearly independent. We will use the Gram-Schmidt Process to convert this to an orthogonal basis, $\{\vec{w}_1, \vec{w}_2\}$. In this case, as we remarked it is only necessary to find an orthogonal basis, and it is not required that it be orthonormal.

$$\vec{w}_1 = \vec{u}_1 = \left[\begin{array}{r} 1 \\ 0 \\ -1 \end{array}\right]$$

$$\begin{aligned}
\vec{w}_2 &= \vec{u}_2 - \left(\frac{\vec{u}_2 \bullet \vec{w}_1}{\|\vec{w}_1\|^2}\right)\vec{w}_1 \\
&= \left[\begin{array}{c} 0 \\ 1 \\ 2 \end{array}\right] - \left(\frac{-2}{2}\right)\left[\begin{array}{r} 1 \\ 0 \\ -1 \end{array}\right] \\
&= \left[\begin{array}{c} 0 \\ 1 \\ 2 \end{array}\right] + \left[\begin{array}{r} 1 \\ 0 \\ -1 \end{array}\right] \\
&= \left[\begin{array}{c} 1 \\ 1 \\ 1 \end{array}\right]
\end{aligned}$$

Therefore an orthogonal basis of W is

$$\{\vec{w}_1, \vec{w}_2\} = \left\{ \left[\begin{array}{r} 1 \\ 0 \\ -1 \end{array}\right], \left[\begin{array}{c} 1 \\ 1 \\ 1 \end{array}\right] \right\}$$

We can now use this basis to find the orthogonal projection of the point $Y = (1, 0, 3)$ on the subspace W. We will write the position vector $\vec{y}$ of Y as $\vec{y} = \left[\begin{array}{c} 1 \\ 0 \\ 3 \end{array}\right]$. Using Definition 4.136, we compute the projection as follows:

$$\begin{aligned}
\vec{z} &= \mathrm{proj}_W(\vec{y}) \\
&= \left(\frac{\vec{y} \bullet \vec{w}_1}{\|\vec{w}_1\|^2}\right)\vec{w}_1 + \left(\frac{\vec{y} \bullet \vec{w}_2}{\|\vec{w}_2\|^2}\right)\vec{w}_2 \\
&= \left(\frac{-2}{2}\right)\left[\begin{array}{r} 1 \\ 0 \\ -1 \end{array}\right] + \left(\frac{4}{3}\right)\left[\begin{array}{c} 1 \\ 1 \\ 1 \end{array}\right] \\
&= \left[\begin{array}{c} \frac{1}{3} \\ \frac{4}{3} \\ \frac{7}{3} \end{array}\right]
\end{aligned}$$

Therefore the point Z on W closest to the point $(1, 0, 3)$ is $\left(\frac{1}{3}, \frac{4}{3}, \frac{7}{3}\right)$.

$\square$

Recall that the vector $\vec{y} - \vec{z}$ is perpendicular (orthogonal) to all the vectors contained in the plane W. Using a basis for W, we can in fact find all such vectors which are perpendicular to W. We call this set of vectors the **orthogonal complement** of W and denote it $W^{\perp}$.

Definition 4.139: Orthogonal Complement

Let W be a subspace of $\mathbb{R}^n$. Then the orthogonal complement of W, written $W^{\perp}$, is the set of all vectors $\vec{x}$ such that $\vec{x} \bullet \vec{z} = 0$ for all vectors $\vec{z}$ in W.

$$W^{\perp} = \{\vec{x} \in \mathbb{R}^n \text{ such that } \vec{x} \bullet \vec{z} = 0 \text{ for all } \vec{z} \in W\}$$

The orthogonal complement is defined as the set of all vectors which are orthogonal to all vectors in the original subspace. It turns out that it is sufficient that the vectors in the orthogonal complement be orthogonal to a spanning set of the original space.

Proposition 4.140: Orthogonal to Spanning Set

Let W be a subspace of $\mathbb{R}^n$ such that $W = \text{span}\{\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_m\}$. Then $W^{\perp}$ is the set of all vectors which are orthogonal to each $\vec{w}_i$ in the spanning set.

The following proposition demonstrates that the orthogonal complement of a subspace is itself a subspace.

Proposition 4.141: The Orthogonal Complement

Let W be a subspace of $\mathbb{R}^n$. Then the orthogonal complement $W^{\perp}$ is also a subspace of $\mathbb{R}^n$.

Consider the following proposition.

Proposition 4.142: Orthogonal Complement of $\mathbb{R}^n$

The complement of $\mathbb{R}^n$ is the set containing the zero vector:

$$(\mathbb{R}^n)^{\perp} = \left\{\vec{0}\right\}$$

Similarly,

$$\left\{\vec{0}\right\}^{\perp} = (\mathbb{R}^n)$$

Proof. Here, $\vec{0}$ is the zero vector of $\mathbb{R}^n$. Since $\vec{x} \bullet \vec{0} = 0$ for all $\vec{x} \in \mathbb{R}^n$, $\mathbb{R}^n \subseteq \{\vec{0}\}^{\perp}$. Since $\{\vec{0}\}^{\perp} \subseteq \mathbb{R}^n$, the equality follows, i.e., $\{\vec{0}\}^{\perp} = \mathbb{R}^n$.

Again, since $\vec{x} \bullet \vec{0} = 0$ for all $\vec{x} \in \mathbb{R}^n$, $\vec{0} \in (\mathbb{R}^n)^{\perp}$, so $\{\vec{0}\} \subseteq (\mathbb{R}^n)^{\perp}$. Suppose $\vec{x} \in \mathbb{R}^n$, $\vec{x} \neq \vec{0}$. Since $\vec{x} \bullet \vec{x} = ||\vec{x}||^2$ and $\vec{x} \neq \vec{0}$, $\vec{x} \bullet \vec{x} \neq 0$, so $\vec{x} \notin (\mathbb{R}^n)^{\perp}$. Therefore $(\mathbb{R}^n)^{\perp} \subseteq \{\vec{0}\}$, and thus $(\mathbb{R}^n)^{\perp} = \{\vec{0}\}$. $\square$

In the next example, we will look at how to find $W^{\perp}$.

Example 4.143: Orthogonal Complement

Let W be the plane through the origin given by the equation $x - 2y + z = 0$. Find a basis for the orthogonal complement of W.

Solution.

From Example 4.138 we know that we can write W as

$$W = \text{span}\{\vec{u}_1, \vec{u}_2\} = \text{span}\left\{ \left[\begin{array}{r} 1 \\ 0 \\ -1 \end{array} \right], \left[\begin{array}{r} 0 \\ 1 \\ 2 \end{array} \right] \right\}$$

In order to find $W^\perp$, we need to find all $\vec{x}$ which are orthogonal to every vector in this span.

Let $\vec{x} = \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} \right]$. In order to satisfy $\vec{x} \bullet \vec{u}_1 = 0$, the following equation must hold.

$$x_1 - x_3 = 0$$

In order to satisfy $\vec{x} \bullet \vec{u}_2 = 0$, the following equation must hold.

$$x_2 + 2x_3 = 0$$

Both of these equations must be satisfied, so we have the following system of equations.

$$\begin{array}{c} x_1 - x_3 = 0 \\ x_2 + 2x_3 = 0 \end{array}$$

To solve, set up the augmented matrix.

$$\left[\begin{array}{rrr|r} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \end{array} \right]$$

Using Gaussian Elimination, we find that $W^\perp = \text{span}\left\{ \left[\begin{array}{r} 1 \\ -2 \\ 1 \end{array} \right] \right\}$, and hence $\left\{ \left[\begin{array}{r} 1 \\ -2 \\ 1 \end{array} \right] \right\}$ is a basis for $W^\perp$. ♢

The following results summarize the important properties of the orthogonal projection.

Theorem 4.144: Orthogonal Projection

Let W be a subspace of $\mathbb{R}^n$, Y be any point in $\mathbb{R}^n$, and let Z be the point in W closest to Y. Then,

1. *The position vector $\vec{z}$ of the point Z is given by $\vec{z} = \text{proj}_W(\vec{y})$*
2. *$\vec{z} \in W$ and $\vec{y} - \vec{z} \in W^\perp$*
3. *$|Y - Z| < |Y - Z_1|$ for all $Z_1 \neq Z \in W$*

Consider the following example of this concept.

Example 4.145: Find a Vector Closest to a Given Vector

Let

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \vec{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \vec{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \text{ and } \vec{v} = \begin{bmatrix} 4 \\ 3 \\ -2 \\ 5 \end{bmatrix}.$$

We want to find the vector in $W = span\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$ closest to $\vec{y}$.

Solution. We will first use the Gram-Schmidt Process to construct the orthogonal basis, B, of W:

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix} \right\}.$$

By Theorem 4.144,

$$\text{proj}_U(\vec{v}) = \frac{2}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \frac{5}{1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \frac{12}{6} \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ -1 \\ 5 \end{bmatrix}$$

is the vector in U closest to $\vec{y}$. ☐

Consider the next example.

Example 4.146: Vector Written as a Sum of Two Vectors

Let W be a subspace given by $W = span \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 2 \end{bmatrix} \right\}$, and $Y = (1,2,3,4)$.

Find the point Z in W closest to Y, and moreover write $\vec{y}$ as the sum of a vector in W and a vector in $W^{\perp}$.

Solution. From Theorem 4.137, the point Z in W closest to Y is given by $\vec{z} = \text{proj}_W(\vec{y})$.

Notice that since the above vectors already give an orthogonal basis for W, we have:

$$\begin{aligned}
\vec{z} &= \mathrm{proj}_W(\vec{y}) \\
&= \left(\frac{\vec{y} \bullet \vec{w}_1}{\|\vec{w}_1\|^2}\right)\vec{w}_1 + \left(\frac{\vec{y} \bullet \vec{w}_2}{\|\vec{w}_2\|^2}\right)\vec{w}_2 \\
&= \left(\frac{4}{2}\right)\left[\begin{array}{c} 1 \\ 0 \\ 1 \\ 0 \end{array}\right] + \left(\frac{10}{5}\right)\left[\begin{array}{c} 0 \\ 1 \\ 0 \\ 2 \end{array}\right] \\
&= \left[\begin{array}{c} 2 \\ 2 \\ 2 \\ 4 \end{array}\right]
\end{aligned}$$

Therefore the point in W closest to Y is $Z = (2,2,2,4)$.

Now, we need to write $\vec{y}$ as the sum of a vector in W and a vector in $W^\perp$. This can easily be done as follows:

$$\vec{y} = \vec{z} + (\vec{y} - \vec{z})$$

since $\vec{z}$ is in W and as we have seen $\vec{y} - \vec{z}$ is in $W^\perp$.
The vector $\vec{y} - \vec{z}$ is given by

$$\vec{y} - \vec{z} = \left[\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array}\right] - \left[\begin{array}{c} 2 \\ 2 \\ 2 \\ 4 \end{array}\right] = \left[\begin{array}{r} -1 \\ 0 \\ 1 \\ 0 \end{array}\right]$$

Therefore, we can write $\vec{y}$ as

$$\left[\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array}\right] = \left[\begin{array}{c} 2 \\ 2 \\ 2 \\ 4 \end{array}\right] + \left[\begin{array}{r} -1 \\ 0 \\ 1 \\ 0 \end{array}\right]$$

$\square$

Example 4.147: Point in a Plane Closest to a Given Point

Find the point Z in the plane $3x + y - 2z = 0$ that is closest to the point $Y = (1,1,1)$.

Solution. The solution will proceed as follows.

1. Find a basis X of the subspace W of $\mathbb{R}^3$ defined by the equation $3x + y - 2z = 0$.

2. Orthogonalize the basis X to get an orthogonal basis B of W.

3. Find the projection on W of the position vector of the point Y.

We now begin the solution.

1. $3x+y-2z=0$ is a system of one equation in three variables. Putting the augmented matrix in reduced row-echelon form:
$$\left[\begin{array}{rrr|r} 3 & 1 & -2 & 0 \end{array}\right] \rightarrow \left[\begin{array}{rrr|r} 1 & \frac{1}{3} & -\frac{2}{3} & 0 \end{array}\right]$$
gives general solution $x=\frac{1}{3}s+\frac{2}{3}t$, $y=s$, $z=t$ for any $s,t\in\mathbb{R}$. Then
$$W=\text{span}\left\{\left[\begin{array}{r} -\frac{1}{3} \\ 1 \\ 0 \end{array}\right], \left[\begin{array}{r} \frac{2}{3} \\ 0 \\ 1 \end{array}\right]\right\}$$
Let $X=\left\{\left[\begin{array}{r} -1 \\ 3 \\ 0 \end{array}\right], \left[\begin{array}{r} 2 \\ 0 \\ 3 \end{array}\right]\right\}$. Then X is linearly independent and $\text{span}(X)=W$, so X is a basis of W.

2. Use the Gram-Schmidt Process to get an orthogonal basis of W:
$$\vec{f}_1=\left[\begin{array}{r} -1 \\ 3 \\ 0 \end{array}\right] \text{ and } \vec{f}_2=\left[\begin{array}{r} 2 \\ 0 \\ 3 \end{array}\right]-\frac{-2}{10}\left[\begin{array}{r} -1 \\ 3 \\ 0 \end{array}\right]=\frac{1}{5}\left[\begin{array}{r} 9 \\ 3 \\ 15 \end{array}\right].$$
Therefore $B=\left\{\left[\begin{array}{r} -1 \\ 3 \\ 0 \end{array}\right], \left[\begin{array}{r} 3 \\ 1 \\ 5 \end{array}\right]\right\}$ is an orthogonal basis of W.

3. To find the point Z on W closest to $Y=(1,1,1)$, compute
$$\begin{aligned} \text{proj}_W\left[\begin{array}{r} 1 \\ 1 \\ 1 \end{array}\right] &= \frac{2}{10}\left[\begin{array}{r} -1 \\ 3 \\ 0 \end{array}\right]+\frac{9}{35}\left[\begin{array}{r} 3 \\ 1 \\ 5 \end{array}\right] \\ &= \frac{1}{7}\left[\begin{array}{r} 4 \\ 6 \\ 9 \end{array}\right]. \end{aligned}$$
Therefore, $Z=\left(\frac{4}{7},\frac{6}{7},\frac{9}{7}\right)$.

$\square$

Exercises

Exercise 4.11.1 *Determine whether the following set of vectors is orthogonal. If it is orthogonal, determine whether it is also orthonormal.*
$$\left[\begin{array}{c} \frac{1}{6}\sqrt{2}\sqrt{3} \\ \frac{1}{3}\sqrt{2}\sqrt{3} \\ -\frac{1}{6}\sqrt{2}\sqrt{3} \end{array}\right], \left[\begin{array}{c} \frac{1}{2}\sqrt{2} \\ 0 \\ \frac{1}{2}\sqrt{2} \end{array}\right], \left[\begin{array}{c} -\frac{1}{3}\sqrt{3} \\ \frac{1}{3}\sqrt{3} \\ \frac{1}{3}\sqrt{3} \end{array}\right]$$

If the set of vectors is orthogonal but not orthonormal, give an orthonormal set of vectors which has the same span.

Exercise 4.11.2 *Determine whether the following set of vectors is orthogonal. If it is orthogonal, determine whether it is also orthonormal.*

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

If the set of vectors is orthogonal but not orthonormal, give an orthonormal set of vectors which has the same span.

Exercise 4.11.3 *Determine whether the following set of vectors is orthogonal. If it is orthogonal, determine whether it is also orthonormal.*

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

If the set of vectors is orthogonal but not orthonormal, give an orthonormal set of vectors which has the same span.

Exercise 4.11.4 *Determine whether the following set of vectors is orthogonal. If it is orthogonal, determine whether it is also orthonormal.*

$$\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

If the set of vectors is orthogonal but not orthonormal, give an orthonormal set of vectors which has the same span.

Exercise 4.11.5 *Determine whether the following set of vectors is orthogonal. If it is orthogonal, determine whether it is also orthonormal.*

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

If the set of vectors is orthogonal but not orthonormal, give an orthonormal set of vectors which has the same span.

Exercise 4.11.6 *Here are some matrices. Label according to whether they are symmetric, skew symmetric, or orthogonal.*

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

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

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

Exercise 4.11.7 *For U an orthogonal matrix, explain why $\|U\vec{x}\| = \|\vec{x}\|$ for any vector $\vec{x}$. Next explain why if U is an $n \times n$ matrix with the property that $\|U\vec{x}\| = \|\vec{x}\|$ for all vectors, $\vec{x}$, then U must be orthogonal. Thus the orthogonal matrices are exactly those which preserve length.*

Exercise 4.11.8 *Suppose U is an orthogonal $n \times n$ matrix. Explain why rank$(U) = n$.*

Exercise 4.11.9 *Fill in the missing entries to make the matrix orthogonal.*

$$\begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \_ & \_ \\ \_ & \frac{\sqrt{6}}{3} & \_ \end{bmatrix}.$$

Exercise 4.11.10 *Fill in the missing entries to make the matrix orthogonal.*

$$\begin{bmatrix} \frac{2}{3} & \frac{\sqrt{2}}{2} & \frac{1}{6}\sqrt{2} \\ \frac{2}{3} & \_ & \_ \\ \_ & 0 & \_ \end{bmatrix}$$

Exercise 4.11.11 *Fill in the missing entries to make the matrix orthogonal.*

$$\begin{bmatrix} \frac{1}{3} & -\frac{2}{\sqrt{5}} & \_ \\ \frac{2}{3} & 0 & \_ \\ \_ & \_ & \frac{4}{15}\sqrt{5} \end{bmatrix}$$

Exercise 4.11.12 *Find an orthonormal basis for the span of each of the following sets of vectors.*

1. $\begin{bmatrix} 3 \\ -4 \\ 0 \end{bmatrix}, \begin{bmatrix} 7 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 7 \\ 1 \end{bmatrix}$

2. $\begin{bmatrix} 3 \\ 0 \\ -4 \end{bmatrix}, \begin{bmatrix} 11 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 7 \end{bmatrix}$

3. $\begin{bmatrix} 3 \\ 0 \\ -4 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \\ 10 \end{bmatrix}, \begin{bmatrix} -7 \\ 1 \\ 1 \end{bmatrix}$

Exercise 4.11.13 *Using the Gram Schmidt process find an orthonormal basis for the following span:*

$$span \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

Exercise 4.11.14 *Using the Gram Schmidt process find an orthonormal basis for the following span:*

$$span \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Exercise 4.11.15 *The set* $V = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : 2x + 3y - z = 0 \right\}$ *is a subspace of* $\mathbb{R}^3$. *Find an orthonormal basis for this subspace.*

Exercise 4.11.16 *Consider the following scalar equation of a plane.*

$$2x - 3y + z = 0$$

Find the orthogonal complement of the vector $\vec{v} = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}$. *Also find the point on the plane which is closest to* $(3, 4, 1)$.

Exercise 4.11.17 *Consider the following scalar equation of a plane.*

$$x + 3y + z = 0$$

Find the orthogonal complement of the vector $\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$. *Also find the point on the plane which is closest to* $(3, 4, 1)$.

Exercise 4.11.18 *Let* $\vec{v}$ *be a vector and let* $\vec{n}$ *be a normal vector for a plane through the origin. Find the equation of the line through the point determined by* $\vec{v}$ *which has direction vector* $\vec{n}$. *Show that it intersects the plane at the point determined by* $\vec{v} - proj_{\vec{n}}\vec{v}$. ***Hint:*** *The line:* $\vec{v} + t\vec{n}$. *It is in the plane if* $\vec{n} \bullet (\vec{v} + t\vec{n}) = 0$. *Determine* t. *Then substitute in to the equation of the line.*

Exercise 4.11.19 *As shown in the above problem, one can find the closest point to $\vec{v}$ in a plane through the origin by finding the intersection of the line through $\vec{v}$ having direction vector equal to the normal vector to the plane with the plane. If the plane does not pass through the origin, this will still work to find the point on the plane closest to the point determined by $\vec{v}$. Here is a relation which defines a plane*

$$2x + y + z = 11$$

and here is a point: $(1,1,2)$. *Find the point on the plane which is closest to this point. Then determine the distance from the point to the plane by taking the distance between these two points.* ***Hint:*** *Line:* $(x, y, z) = (1, 1, 2) + t\,(2, 1, 1)$. *Now require that it intersect the plane.*

Exercise 4.11.20 *In general, you have a point* (x_0, y_0, z_0) *and a scalar equation for a plane* $ax + by + cz = d$ *where* $a^2 + b^2 + c^2 > 0$. *Determine a formula for the closest point on the plane to the given point. Then use this point to get a formula for the distance from the given point to the plane.* ***Hint:*** *Find the line perpendicular to the plane which goes through the given point:* $(x, y, z) = (x_0, y_0, z_0) + t\,(a, b, c)$. *Now require that this point satisfy the equation for the plane to determine* t.

5. Linear Transformations

5.1 Linear Transformations

Outcomes

A. *Understand the definition of a linear transformation, and that all linear transformations are determined by matrix multiplication.*

Recall that when we multiply an $m \times n$ matrix by an $n \times 1$ column vector, the result is an $m \times 1$ column vector. In this section we will discuss how, through matrix multiplication, an $m \times n$ matrix **transforms** an $n \times 1$ column vector into an $m \times 1$ column vector.

Recall that the $n \times 1$ vector given by

$$\vec{x} = \left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array} \right]$$

is said to belong to $\mathbb{R}^n$, which is the set of all $n \times 1$ vectors. In this section, we will discuss transformations of vectors in $\mathbb{R}^n$.

Consider the following example.

Example 5.1: A Function Which Transforms Vectors

Consider the matrix $A = \left[\begin{array}{ccc} 1 & 2 & 0 \\ 2 & 1 & 0 \end{array} \right]$. *Show that by matrix multiplication* A *transforms vectors in* $\mathbb{R}^3$ *into vectors in* $\mathbb{R}^2$.

Solution. First, recall that vectors in $\mathbb{R}^3$ are vectors of size 3×1, while vectors in $\mathbb{R}^2$ are of size 2×1. If we multiply A, which is a 2×3 matrix, by a 3×1 vector, the result will be a 2×1 vector. This what we mean when we say that A *transforms* vectors.

Now, for $\left[\begin{array}{c} x \\ y \\ z \end{array} \right]$ in $\mathbb{R}^3$, multiply on the left by the given matrix to obtain the new vector. This product looks like

$$\left[\begin{array}{ccc} 1 & 2 & 0 \\ 2 & 1 & 0 \end{array} \right] \left[\begin{array}{c} x \\ y \\ z \end{array} \right] = \left[\begin{array}{c} x + 2y \\ 2x + y \end{array} \right]$$

The resulting product is a 2×1 vector which is determined by the choice of x and y. Here are some numerical examples.

$$\left[\begin{array}{ccc} 1 & 2 & 0 \\ 2 & 1 & 0 \end{array}\right]\left[\begin{array}{c} 1 \\ 2 \\ 3 \end{array}\right]=\left[\begin{array}{c} 5 \\ 4 \end{array}\right]$$

Here, the vector $\left[\begin{array}{c} 1 \\ 2 \\ 3 \end{array}\right]$ in $\mathbb{R}^3$ was transformed by the matrix into the vector $\left[\begin{array}{c} 5 \\ 4 \end{array}\right]$ in $\mathbb{R}^2$.

Here is another example:

$$\left[\begin{array}{ccc} 1 & 2 & 0 \\ 2 & 1 & 0 \end{array}\right]\left[\begin{array}{r} 10 \\ 5 \\ -3 \end{array}\right]=\left[\begin{array}{c} 20 \\ 25 \end{array}\right]$$

♠

The idea is to define a function which takes vectors in $\mathbb{R}^3$ and delivers new vectors in $\mathbb{R}^2$. In this case, that function is multiplication by the matrix A.

Let T denote such a function. The notation $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ means that the function T transforms vectors in $\mathbb{R}^n$ into vectors in $\mathbb{R}^m$. The notation $T(\vec{x})$ means the transformation T applied to the vector $\vec{x}$. The above example demonstrated a transformation achieved by matrix multiplication. In this case, we often write

$$T_A(\vec{x}) = A\vec{x}$$

Therefore, T_A is the transformation determined by the matrix A. In this case we say that T is a matrix transformation.

Recall the property of matrix multiplication that states that for k and p scalars,

$$A(kB + pC) = kAB + pAC$$

In particular, for A an $m \times n$ matrix and B and C, $n \times 1$ vectors in $\mathbb{R}^n$, this formula holds.

In other words, this means that matrix multiplication gives an example of a linear transformation, which we will now define.

Definition 5.2: Linear Transformation

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a function, where for each $\vec{x} \in \mathbb{R}^n, T(\vec{x}) \in \mathbb{R}^m$. Then T is a ***linear transformation*** *if whenever k, p are scalars and $\vec{x}_1$ and $\vec{x}_2$ are vectors in $\mathbb{R}^n$ ($n \times 1$ vectors),*

$$T(k\vec{x}_1 + p\vec{x}_2) = kT(\vec{x}_1) + pT(\vec{x}_2)$$

Consider the following example.

Example 5.3: Linear Transformation

Let T be a transformation defined by $T : \mathbb{R}^3 \to \mathbb{R}^2$ is defined by

$$T \left[\begin{array}{c} x \\ y \\ z \end{array} \right] = \left[\begin{array}{c} x+y \\ x-z \end{array} \right] \text{ for all } \left[\begin{array}{c} x \\ y \\ z \end{array} \right] \in \mathbb{R}^3$$

Show that T is a linear transformation.

Solution. By Definition 5.2 we need to show that $T\left(k\vec{x}_1 + p\vec{x}_2\right) = kT\left(\vec{x}_1\right) + pT\left(\vec{x}_2\right)$ for all scalars k, p and vectors $\vec{x}_1, \vec{x}_2$. Let

$$\vec{x}_1 = \left[\begin{array}{c} x_1 \\ y_1 \\ z_1 \end{array} \right], \vec{x}_2 = \left[\begin{array}{c} x_2 \\ y_2 \\ z_2 \end{array} \right]$$

Then

$$\begin{aligned}
T\left(k\vec{x}_1 + p\vec{x}_2\right) &= T\left(k \left[\begin{array}{c} x_1 \\ y_1 \\ z_1 \end{array} \right] + p \left[\begin{array}{c} x_2 \\ y_2 \\ z_2 \end{array} \right] \right) \\
&= T\left(\left[\begin{array}{c} kx_1 \\ ky_1 \\ kz_1 \end{array} \right] + \left[\begin{array}{c} px_2 \\ py_2 \\ pz_2 \end{array} \right] \right) \\
&= T\left(\left[\begin{array}{c} kx_1 + px_2 \\ ky_1 + py_2 \\ kz_1 + pz_2 \end{array} \right] \right) \\
&= \left[\begin{array}{c} (kx_1 + px_2) + (ky_1 + py_2) \\ (kx_1 + px_2) - (kz_1 + pz_2) \end{array} \right] \\
&= \left[\begin{array}{c} (kx_1 + ky_1) + (px_2 + py_2) \\ (kx_1 - kz_1) + (px_2 - pz_2) \end{array} \right] \\
&= \left[\begin{array}{c} kx_1 + ky_1 \\ kx_1 - kz_1 \end{array} \right] + \left[\begin{array}{c} px_2 + py_2 \\ px_2 - pz_2 \end{array} \right] \\
&= k \left[\begin{array}{c} x_1 + y_1 \\ x_1 - z_1 \end{array} \right] + p \left[\begin{array}{c} x_2 + y_2 \\ x_2 - z_2 \end{array} \right] \\
&= kT(\vec{x}_1) + pT(\vec{x}_2)
\end{aligned}$$

Therefore T is a linear transformation. ◻

Two important examples of linear transformations are the zero transformation and identity transformation. The zero transformation defined by $T\left(\vec{x}\right) = \vec{(0)}$ for all $\vec{x}$ is an example of a linear transformation. Similarly the identity transformation defined by $T\left(\vec{x}\right) = \vec{(x)}$ is also linear. Take the time to prove these using the method demonstrated in Example 5.3.

We began this section by discussing matrix transformations, where multiplication by a matrix transforms vectors. These matrix transformations are in fact linear transformations.

Theorem 5.4: Matrix Transformations are Linear Transformations

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a transformation defined by $T(\vec{x}) = A\vec{x}$. Then T is a linear transformation.

It turns out that every linear transformation can be expressed as a matrix transformation, and thus linear transformations are exactly the same as matrix transformations.

Exercises

Exercise 5.1.1 *Show the map $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ defined by $T(\vec{x}) = A\vec{x}$ where A is an $m \times n$ matrix and $\vec{x}$ is an $m \times 1$ column vector is a linear transformation.*

Exercise 5.1.2 *Show that the function $T_{\vec{u}}$ defined by $T_{\vec{u}}(\vec{v}) = \vec{v} - proj_{\vec{u}}(\vec{v})$ is also a linear transformation.*

Exercise 5.1.3 *Let $\vec{u}$ be a fixed vector. The function $T_{\vec{u}}$ defined by $T_{\vec{u}}\vec{v} = \vec{u} + \vec{v}$ has the effect of translating all vectors by adding $\vec{u} \neq \vec{0}$. Show this is not a linear transformation. Explain why it is not possible to represent $T_{\vec{u}}$ in $\mathbb{R}^3$ by multiplying by a 3×3 matrix.*

5.2 The Matrix of a Linear Transformation

Outcomes

A. *Find the matrix of a linear transformation and determine the action on a vector in $\mathbb{R}^n$.*

In the above examples, the action of the linear transformations was to multiply by a matrix. It turns out that this is always the case for linear transformations. If T is **any** linear transformation which maps $\mathbb{R}^n$ to $\mathbb{R}^m$, there is **always** an $m \times n$ matrix A with the property that

$$T(\vec{x}) = A\vec{x} \tag{5.1}$$

for all $\vec{x} \in \mathbb{R}^n$.

Theorem 5.5: Matrix of a Linear Transformation

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear transformation. Then we can find a matrix A such that $T(\vec{x}) = A\vec{x}$. In this case, we say that T is determined *or* induced *by the matrix A.*

Here is why. Suppose $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ is a linear transformation and you want to find the matrix defined by this linear transformation as described in 5.1. Note that

$$\vec{x} = \left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array}\right] = x_1 \left[\begin{array}{c} 1 \\ 0 \\ \vdots \\ 0 \end{array}\right] + x_2 \left[\begin{array}{c} 0 \\ 1 \\ \vdots \\ 0 \end{array}\right] + \cdots + x_n \left[\begin{array}{c} 0 \\ 0 \\ \vdots \\ 1 \end{array}\right] = \sum_{i=1}^{n} x_i \vec{e}_i$$

where $\vec{e}_i$ is the i^{th} column of I_n, that is the $n \times 1$ vector which has zeros in every slot but the i^{th} and a 1 in this slot.

Then since T is linear,

$$\begin{aligned} T(\vec{x}) &= \sum_{i=1}^{n} x_i T(\vec{e}_i) \\ &= \left[\begin{array}{ccc} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & | \end{array}\right] \left[\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array}\right] \\ &= A \left[\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array}\right] \end{aligned}$$

Therefore, the desired matrix is obtained from constructing the i^{th} column as $T(\vec{e}_i)$. We state this formally as the following theorem.

Theorem 5.6: Matrix of a Linear Transformation

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear transformation. Then the matrix A satisfying $T(\vec{x}) = A\vec{x}$ is given by

$$A = \left[\begin{array}{ccc} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & | \end{array}\right]$$

where $\vec{e}_i$ is the i^{th} column of I_n, and then $T(\vec{e}_i)$ is the i^{th} column of A.

The following Corollary is an essential result.

Corollary 5.7: Matrix and Linear Transformation

A transformation T is a linear transformation if and only if it is a matrix transformation.

Consider the following example.

Example 5.8: The Matrix of a Linear Transformation

Suppose T is a linear transformation, $T : \mathbb{R}^3 \to \mathbb{R}^2$ where

$$T\left[\begin{array}{c}1\\0\\0\end{array}\right]=\left[\begin{array}{c}1\\2\end{array}\right],\ T\left[\begin{array}{c}0\\1\\0\end{array}\right]=\left[\begin{array}{r}9\\-3\end{array}\right],\ T\left[\begin{array}{c}0\\0\\1\end{array}\right]=\left[\begin{array}{c}1\\1\end{array}\right]$$

Find the matrix A of T such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x}$.

Solution. By Theorem 5.6 we construct A as follows:

$$A=\left[\begin{array}{ccc}| & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & |\end{array}\right]$$

In this case, A will be a 2×3 matrix, so we need to find $T(\vec{e}_1), T(\vec{e}_2)$, and $T(\vec{e}_3)$. Luckily, we have been given these values so we can fill in A as needed, using these vectors as the columns of A. Hence,

$$A=\left[\begin{array}{rrr}1 & 9 & 1\\2 & -3 & 1\end{array}\right]$$

$\square$

In this example, we were given the resulting vectors of $T(\vec{e}_1), T(\vec{e}_2)$, and $T(\vec{e}_3)$. Constructing the matrix A was simple, as we could simply use these vectors as the columns of A. The next example shows how to find A when we are not given the $T(\vec{e}_i)$ so clearly.

Example 5.9: The Matrix of Linear Transformation: Inconveniently Defined

Suppose T is a linear transformation, $T : \mathbb{R}^2 \to \mathbb{R}^2$ and

$$T\left[\begin{array}{c}1\\1\end{array}\right]=\left[\begin{array}{c}1\\2\end{array}\right],\ T\left[\begin{array}{r}0\\-1\end{array}\right]=\left[\begin{array}{c}3\\2\end{array}\right]$$

Find the matrix A of T such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x}$.

Solution. By Theorem 5.6 to find this matrix, we need to determine the action of T on $\vec{e}_1$ and $\vec{e}_2$. In Example 5.8, we were given these resulting vectors. However, in this example, we have been given T of two different vectors. How can we find out the action of T on $\vec{e}_1$ and $\vec{e}_2$? In particular for $\vec{e}_1$, suppose there exist x and y such that

$$\left[\begin{array}{c}1\\0\end{array}\right]=x\left[\begin{array}{c}1\\1\end{array}\right]+y\left[\begin{array}{r}0\\-1\end{array}\right] \tag{5.2}$$

Then, since T is linear,

$$T\left[\begin{array}{c}1\\0\end{array}\right]=xT\left[\begin{array}{c}1\\1\end{array}\right]+yT\left[\begin{array}{c}0\\-1\end{array}\right]$$

Substituting in values, this sum becomes

$$T\left[\begin{array}{c}1\\0\end{array}\right]=x\left[\begin{array}{c}1\\2\end{array}\right]+y\left[\begin{array}{c}3\\2\end{array}\right] \tag{5.3}$$

Therefore, if we know the values of x and y which satisfy 5.2, we can substitute these into equation 5.3. By doing so, we find $T(\vec{e}_1)$ which is the first column of the matrix A.

We proceed to find x and y. We do so by solving 5.2, which can be done by solving the system

$$\begin{array}{c}x=1\\x-y=0\end{array}$$

We see that $x=1$ and $y=1$ is the solution to this system. Substituting these values into equation 5.3, we have

$$T\left[\begin{array}{c}1\\0\end{array}\right]=1\left[\begin{array}{c}1\\2\end{array}\right]+1\left[\begin{array}{c}3\\2\end{array}\right]=\left[\begin{array}{c}1\\2\end{array}\right]+\left[\begin{array}{c}3\\2\end{array}\right]=\left[\begin{array}{c}4\\4\end{array}\right]$$

Therefore $\left[\begin{array}{c}4\\4\end{array}\right]$ is the first column of A.

Computing the second column is done in the same way, and is left as an exercise.

The resulting matrix A is given by

$$A=\left[\begin{array}{cc}4&-3\\4&-2\end{array}\right]$$

□

This example illustrates a very long procedure for finding the matrix of A. While this method is reliable and will always result in the correct matrix A, the following procedure provides an alternative method.

Procedure 5.10: Finding the Matrix of Inconveniently Defined Linear Transformation

Suppose $T:\mathbb{R}^n\to\mathbb{R}^m$ is a linear transformation. Suppose there exist vectors $\{\vec{a}_1,\cdots,\vec{a}_n\}$ in $\mathbb{R}^n$ such that $\left[\begin{array}{ccc}\vec{a}_1&\cdots&\vec{a}_n\end{array}\right]^{-1}$ exists, and

$$T(\vec{a}_i)=\vec{b}_i$$

Then the matrix of T must be of the form

$$\left[\begin{array}{ccc}\vec{b}_1&\cdots&\vec{b}_n\end{array}\right]\left[\begin{array}{ccc}\vec{a}_1&\cdots&\vec{a}_n\end{array}\right]^{-1}$$

We will illustrate this procedure in the following example. You may also find it useful to work through Example 5.9 using this procedure.

Example 5.11: Matrix of a Linear Transformation Given Inconveniently

Suppose $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear transformation and

$$T \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, T \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, T \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Find the matrix of this linear transformation.

Solution. By Procedure 5.10, $A = \begin{bmatrix} 1 & 0 & 1 \\ 3 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}^{-1}$ and $B = \begin{bmatrix} 0 & 2 & 0 \\ 1 & 1 & 0 \\ 1 & 3 & 1 \end{bmatrix}$

Then, Procedure 5.10 claims that the matrix of T is

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

Indeed you can first verify that $T(\vec{x}) = C\vec{x}$ for the 3 vectors above:

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

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

But more generally $T(\vec{x}) = C\vec{x}$ for any $\vec{x}$. To see this, let $\vec{y} = A^{-1}\vec{x}$ and then using linearity of T:

$$T(\vec{x}) = T(A\vec{y}) = T\left(\sum_i \vec{y}_i \vec{a}_i\right) = \sum \vec{y}_i T(\vec{a}_i) \sum \vec{y}_i \vec{b}_i = B\vec{y} = BA^{-1}\vec{x} = C\vec{x}$$

$\square$

Recall the dot product discussed earlier. Consider the map $\vec{v} \mapsto \operatorname{proj}_{\vec{u}}(\vec{v})$ which takes a vector a transforms it to its projection onto a given vector $\vec{u}$. It turns out that this map is linear, a result which follows from the properties of the dot product. This is shown as follows.

$$\begin{aligned} \operatorname{proj}_{\vec{u}}(k\vec{v} + p\vec{w}) &= \left(\frac{(k\vec{v} + p\vec{w}) \bullet \vec{u}}{\vec{u} \bullet \vec{u}}\right)\vec{u} \\ &= k\left(\frac{\vec{v} \bullet \vec{u}}{\vec{u} \bullet \vec{u}}\right)\vec{u} + p\left(\frac{\vec{w} \bullet \vec{u}}{\vec{u} \bullet \vec{u}}\right)\vec{u} \\ &= k \operatorname{proj}_{\vec{u}}(\vec{v}) + p \operatorname{proj}_{\vec{u}}(\vec{w}) \end{aligned}$$

Consider the following example.

Example 5.12: Matrix of a Projection Map

Let $\vec{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ *and let* T *be the projection map* $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$ *defined by*

$$T(\vec{v}) = proj_{\vec{u}}(\vec{v})$$

for any $\vec{v} \in \mathbb{R}^3$.

1. *Does this transformation come from multiplication by a matrix?*
2. *If so, what is the matrix?*

Solution.

1. First, we have just seen that $T(\vec{v}) = \mathrm{proj}_{\vec{u}}(\vec{v})$ is linear. Therefore by Theorem 5.5, we can find a matrix A such that $T(\vec{x}) = A\vec{x}$.

2. The columns of the matrix for T are defined above as $T(\vec{e}_i)$. It follows that $T(\vec{e}_i) = \mathrm{proj}_{\vec{u}}(\vec{e}_i)$ gives the i^{th} column of the desired matrix. Therefore, we need to find

$$\mathrm{proj}_{\vec{u}}(\vec{e}_i) = \left(\frac{\vec{e}_i \bullet \vec{u}}{\vec{u} \bullet \vec{u}} \right) \vec{u}$$

For the given vector $\vec{u}$, this implies the columns of the desired matrix are

$$\frac{1}{14} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \frac{2}{14} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \frac{3}{14} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

which you can verify. Hence the matrix of T is

$$\frac{1}{14} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

□

Exercises

Exercise 5.2.1 *Consider the following functions which map* $\mathbb{R}^n$ *to* $\mathbb{R}^n$.

1. *T multiplies the j^{th} component of $\vec{x}$ by a nonzero number b.*
2. *T replaces the i^{th} component of $\vec{x}$ with b times the j^{th} component added to the i^{th} component.*
3. *T switches the i^{th} and j^{th} components.*

Show these functions are linear transformations and describe their matrices A such that $T(\vec{x}) = A\vec{x}$.

Exercise 5.2.2 *You are given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ and you know that*

$$T(A_i) = B_i$$

where $\begin{bmatrix} A_1 & \cdots & A_n \end{bmatrix}^{-1}$ exists. Show that the matrix of T is of the form

$$\begin{bmatrix} B_1 & \cdots & B_n \end{bmatrix} \begin{bmatrix} A_1 & \cdots & A_n \end{bmatrix}^{-1}$$

Exercise 5.2.3 *Suppose T is a linear transformation such that*

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

Find the matrix of T. That is find A such that $T(\vec{x}) = A\vec{x}$.

Exercise 5.2.4 *Suppose T is a linear transformation such that*

$$T \begin{bmatrix} 1 \\ 1 \\ -8 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$$
$$T \begin{bmatrix} -1 \\ 0 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$$
$$T \begin{bmatrix} 0 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \\ -1 \end{bmatrix}$$

Find the matrix of T. That is find A such that $T(\vec{x}) = A\vec{x}$.

Exercise 5.2.5 *Suppose T is a linear transformation such that*

$$T\begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \\ 3 \end{bmatrix}$$
$$T\begin{bmatrix} -1 \\ -2 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ -3 \end{bmatrix}$$
$$T\begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ -3 \end{bmatrix}$$

Find the matrix of T. That is find A such that $T(\vec{x}) = A\vec{x}$.

Exercise 5.2.6 *Suppose T is a linear transformation such that*

$$T\begin{bmatrix} 1 \\ 1 \\ -7 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$$
$$T\begin{bmatrix} -1 \\ 0 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$
$$T\begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}$$

Find the matrix of T. That is find A such that $T(\vec{x}) = A\vec{x}$.

Exercise 5.2.7 *Suppose T is a linear transformation such that*

$$T\begin{bmatrix} 1 \\ 2 \\ -18 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 5 \end{bmatrix}$$
$$T\begin{bmatrix} -1 \\ -1 \\ 15 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 5 \end{bmatrix}$$
$$T\begin{bmatrix} 0 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ -2 \end{bmatrix}$$

Find the matrix of T. That is find A such that $T(\vec{x}) = A\vec{x}$.

Exercise 5.2.8 *Consider the following functions $T : \mathbb{R}^3 \to \mathbb{R}^2$. Show that each is a linear transformation and determine for each the matrix A such that $T(\vec{x}) = A\vec{x}$.*

1. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+2y+3z \\ 2y-3x+z \end{bmatrix}$

2. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 7x+2y+z \\ 3x-11y+2z \end{bmatrix}$

3. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3x+2y+z \\ x+2y+6z \end{bmatrix}$

4. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2y-5x+z \\ x+y+z \end{bmatrix}$

Exercise 5.2.9 *Consider the following functions $T : \mathbb{R}^3 \to \mathbb{R}^2$. Explain why each of these functions T is not linear.*

1. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+2y+3z+1 \\ 2y-3x+z \end{bmatrix}$

2. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+2y^2+3z \\ 2y+3x+z \end{bmatrix}$

3. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \sin x+2y+3z \\ 2y+3x+z \end{bmatrix}$

4. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+2y+3z \\ 2y+3x-\ln z \end{bmatrix}$

Exercise 5.2.10 *Suppose*

$$\begin{bmatrix} A_1 & \cdots & A_n \end{bmatrix}^{-1}$$

*exists where each $A_j \in \mathbb{R}^n$ and let vectors $\{B_1, \cdots, B_n\}$ in $\mathbb{R}^m$ be given. Show that there **always** exists a linear transformation T such that $T(A_i) = B_i$.*

Exercise 5.2.11 *Find the matrix for $T(\vec{w}) = proj_{\vec{v}}(\vec{w})$ where $\vec{v} = \begin{bmatrix} 1 & -2 & 3 \end{bmatrix}^T$.*

Exercise 5.2.12 *Find the matrix for $T(\vec{w}) = proj_{\vec{v}}(\vec{w})$ where $\vec{v} = \begin{bmatrix} 1 & 5 & 3 \end{bmatrix}^T$.*

Exercise 5.2.13 *Find the matrix for $T(\vec{w}) = proj_{\vec{v}}(\vec{w})$ where $\vec{v} = \begin{bmatrix} 1 & 0 & 3 \end{bmatrix}^T$.*

5.3 Properties of Linear Transformations

Outcomes

A. *Use properties of linear transformations to solve problems.*

B. *Find the composite of transformations and the inverse of a transformation.*

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear transformation. Then there are some important properties of T which will be examined in this section. Consider the following theorem.

Theorem 5.13: Properties of Linear Transformations

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear transformation and let $\vec{x} \in \mathbb{R}^n$.

- *T preserves the zero vector.*

$$T(0\vec{x}) = 0T(\vec{x}). \text{ Hence } T(\vec{0}) = \vec{0}$$

- *T preserves the negative of a vector:*

$$T((-1)\vec{x}) = (-1)T(\vec{x}). \text{ Hence } T(-\vec{x}) = -T(\vec{x}).$$

- *T preserves linear combinations:*

$$\text{Let } \vec{x}_1, ..., \vec{x}_k \in \mathbb{R}^n \text{ and } a_1, ..., a_k \in \mathbb{R}.$$

$$\text{Then if } \vec{y} = a_1\vec{x}_1 + a_2\vec{x}_2 + ... + a_k\vec{x}_k, \text{ it follows that}$$
$$T(\vec{y}) = T(a_1\vec{x}_1 + a_2\vec{x}_2 + ... + a_k\vec{x}_k) = a_1T(\vec{x}_1) + a_2T(\vec{x}_2) + ... + a_kT(\vec{x}_k).$$

These properties are useful in determining the action of a transformation on a given vector. Consider the following example.

Example 5.14: Linear Combination

Let $T : \mathbb{R}^3 \mapsto \mathbb{R}^4$ be a linear transformation such that

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

Find $T \begin{bmatrix} -7 \\ 3 \\ -9 \end{bmatrix}$.

Solution. Using the third property in Theorem 5.13, we can find $T \begin{bmatrix} -7 \\ 3 \\ -9 \end{bmatrix}$ by writing $\begin{bmatrix} -7 \\ 3 \\ -9 \end{bmatrix}$ as a linear combination of $\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 4 \\ 0 \\ 5 \end{bmatrix}$.

Therefore we want to find $a, b \in \mathbb{R}$ such that

$$\begin{bmatrix} -7 \\ 3 \\ -9 \end{bmatrix} = a \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} + b \begin{bmatrix} 4 \\ 0 \\ 5 \end{bmatrix}$$

The necessary augmented matrix and resulting reduced row-echelon form are given by:

$$\left[\begin{array}{cc|c} 1 & 4 & -7 \\ 3 & 0 & 3 \\ 1 & 5 & -9 \end{array}\right] \rightarrow \cdots \rightarrow \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{array}\right]$$

Hence $a = 1, b = -2$ and

$$\begin{bmatrix} -7 \\ 3 \\ -9 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} + (-2) \begin{bmatrix} 4 \\ 0 \\ 5 \end{bmatrix}$$

Now, using the third property above, we have

$$\begin{aligned}
T\left[\begin{array}{r} -7 \\ 3 \\ -9 \end{array}\right] &= T\left(1\left[\begin{array}{r} 1 \\ 3 \\ 1 \end{array}\right] + (-2)\left[\begin{array}{r} 4 \\ 0 \\ 5 \end{array}\right]\right) \\
&= 1T\left[\begin{array}{r} 1 \\ 3 \\ 1 \end{array}\right] - 2T\left[\begin{array}{r} 4 \\ 0 \\ 5 \end{array}\right] \\
&= \left[\begin{array}{r} 4 \\ 4 \\ 0 \\ -2 \end{array}\right] - 2\left[\begin{array}{r} 4 \\ 5 \\ -1 \\ 5 \end{array}\right] \\
&= \left[\begin{array}{r} -4 \\ -6 \\ 2 \\ -12 \end{array}\right]
\end{aligned}$$

Therefore, $T\left[\begin{array}{r} -7 \\ 3 \\ -9 \end{array}\right] = \left[\begin{array}{r} -4 \\ -6 \\ 2 \\ -12 \end{array}\right]$. ♠

Suppose two linear transformations act in the same way on $\vec{x}$ for all vectors. Then we say that these transformations are equal.

Definition 5.15: Equal Transformations

Let S and T be linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then $S = T$ if and only if for every $\vec{x} \in \mathbb{R}^n$,

$$S(\vec{x}) = T(\vec{x})$$

Suppose two linear transformations act on the same vector $\vec{x}$, first the transformation T and then a second transformation given by S. We can find the **composite** transformation that results from applying both transformations.

Definition 5.16: Composition of Linear Transformations

*Let $T : \mathbb{R}^k \mapsto \mathbb{R}^n$ and $S : \mathbb{R}^n \mapsto \mathbb{R}^m$ be linear transformations. Then the **composite** of S and T is*

$$S \circ T : \mathbb{R}^k \mapsto \mathbb{R}^m$$

The action of $S \circ T$ is given by

$$(S \circ T)(\vec{x}) = S(T(\vec{x})) \text{ for all } \vec{x} \in \mathbb{R}^k$$

Notice that the resulting vector will be in $\mathbb{R}^m$. Be careful to observe the order of transformations. We write $S \circ T$ but apply the transformation T first, followed by S.

Theorem 5.17: Composition of Transformations

Let $T : \mathbb{R}^k \mapsto \mathbb{R}^n$ and $S : \mathbb{R}^n \mapsto \mathbb{R}^m$ be linear transformations such that T is induced by the matrix A and S is induced by the matrix B. Then $S \circ T$ is a linear transformation which is induced by the matrix BA.

Consider the following example.

Example 5.18: Composition of Transformations

Let T be a linear transformation induced by the matrix

$$A = \left[\begin{array}{cc} 1 & 2 \\ 2 & 0 \end{array} \right]$$

and S a linear transformation induced by the matrix

$$B = \left[\begin{array}{cc} 2 & 3 \\ 0 & 1 \end{array} \right]$$

Find the matrix of the composite transformation $S \circ T$. Then, find $(S \circ T)(\vec{x})$ for $\vec{x} = \left[\begin{array}{c} 1 \\ 4 \end{array} \right]$.

Solution. By Theorem 5.17, the matrix of $S \circ T$ is given by BA.

$$BA = \left[\begin{array}{cc} 2 & 3 \\ 0 & 1 \end{array} \right] \left[\begin{array}{cc} 1 & 2 \\ 2 & 0 \end{array} \right] = \left[\begin{array}{cc} 8 & 4 \\ 2 & 0 \end{array} \right]$$

To find $(S \circ T)(\vec{x})$, multiply $\vec{x}$ by BA as follows

$$\left[\begin{array}{cc} 8 & 4 \\ 2 & 0 \end{array} \right] \left[\begin{array}{c} 1 \\ 4 \end{array} \right] = \left[\begin{array}{r} 24 \\ 2 \end{array} \right]$$

To check, first determine $T(\vec{x})$:

$$\left[\begin{array}{cc} 1 & 2 \\ 2 & 0 \end{array} \right] \left[\begin{array}{c} 1 \\ 4 \end{array} \right] = \left[\begin{array}{c} 9 \\ 2 \end{array} \right]$$

Then, compute $S(T(\vec{x}))$ as follows:

$$\left[\begin{array}{cc} 2 & 3 \\ 0 & 1 \end{array} \right] \left[\begin{array}{c} 9 \\ 2 \end{array} \right] = \left[\begin{array}{r} 24 \\ 2 \end{array} \right]$$

$\square$

Consider a composite transformation $S \circ T$, and suppose that this transformation acted such that $(S \circ T)(\vec{x}) = \vec{x}$. That is, the transformation S took the vector $T(\vec{x})$ and returned it to $\vec{x}$. In this case, S and T are inverses of each other. Consider the following definition.

Definition 5.19: Inverse of a Transformation

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ and $S : \mathbb{R}^n \mapsto \mathbb{R}^n$ be linear transformations. Suppose that for each $\vec{x} \in \mathbb{R}^n$,

$$(S \circ T)(\vec{x}) = \vec{x}$$

and

$$(T \circ S)(\vec{x}) = \vec{x}$$

Then, S is called an inverse of T and T is called an inverse of S. Geometrically, they reverse the action of each other.

The following theorem is crucial, as it claims that the above inverse transformations are unique.

Theorem 5.20: Inverse of a Transformation

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ be a linear transformation induced by the matrix A. Then T has an inverse transformation if and only if the matrix A is invertible. In this case, the inverse transformation is unique and denoted $T^{-1} : \mathbb{R}^n \mapsto \mathbb{R}^n$. T^{-1} is induced by the matrix A^{-1}.

Consider the following example.

Example 5.21: Inverse of a Transformation

Let $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be a linear transformation induced by the matrix

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

Show that T^{-1} exists and find the matrix B which it is induced by.

Solution. Since the matrix A is invertible, it follows that the transformation T is invertible. Therefore, T^{-1} exists.

You can verify that A^{-1} is given by:

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

Therefore the linear transformation T^{-1} is induced by the matrix A^{-1}. ☐

Exercises

Exercise 5.3.1 *Show that if a function $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then it is* ***always*** *the case that $T\left(\vec{0}\right) = \vec{0}$.*

Exercise 5.3.2 *Let T be a linear transformation induced by the matrix $A = \left[\begin{array}{rr} 3 & 1 \\ -1 & 2 \end{array}\right]$ and S a linear transformation induced by $B = \left[\begin{array}{rr} 0 & -2 \\ 4 & 2 \end{array}\right]$. Find matrix of $S \circ T$ and find $(S \circ T)(\vec{x})$ for $\vec{x} = \left[\begin{array}{r} 2 \\ -1 \end{array}\right]$.*

Exercise 5.3.3 *Let T be a linear transformation and suppose $T\left(\left[\begin{array}{r} 1 \\ -4 \end{array}\right]\right) = \left[\begin{array}{r} 2 \\ -3 \end{array}\right]$. Suppose S is a linear transformation induced by the matrix $B = \left[\begin{array}{rr} 1 & 2 \\ -1 & 3 \end{array}\right]$. Find $(S \circ T)(\vec{x})$ for $\vec{x} = \left[\begin{array}{r} 1 \\ -4 \end{array}\right]$.*

Exercise 5.3.4 *Let T be a linear transformation induced by the matrix $A = \left[\begin{array}{rr} 2 & 3 \\ 1 & 1 \end{array}\right]$ and S a linear transformation induced by $B = \left[\begin{array}{rr} -1 & 3 \\ 1 & -2 \end{array}\right]$. Find matrix of $S \circ T$ and find $(S \circ T)(\vec{x})$ for $\vec{x} = \left[\begin{array}{r} 5 \\ 6 \end{array}\right]$.*

Exercise 5.3.5 *Let T be a linear transformation induced by the matrix $A = \left[\begin{array}{rr} 2 & 1 \\ 5 & 2 \end{array}\right]$. Find the matrix of T^{-1}.*

Exercise 5.3.6 *Let T be a linear transformation induced by the matrix $A = \left[\begin{array}{rr} 4 & -3 \\ 2 & -2 \end{array}\right]$. Find the matrix of T^{-1}.*

Exercise 5.3.7 *Let T be a linear transformation and suppose $T\left(\left[\begin{array}{r} 1 \\ 2 \end{array}\right]\right) = \left[\begin{array}{r} 9 \\ 8 \end{array}\right]$, $T\left(\left[\begin{array}{r} 0 \\ -1 \end{array}\right]\right) = \left[\begin{array}{r} -4 \\ -3 \end{array}\right]$. Find the matrix of T^{-1}.*

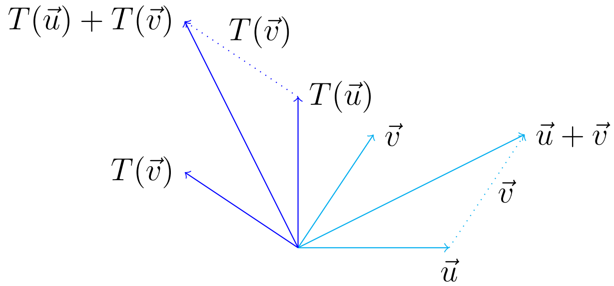
5.4 Special Linear Transformations in $\mathbb{R}^2$

Outcomes

A. *Find the matrix of rotations and reflections in $\mathbb{R}^2$ and determine the action of each on a vector in $\mathbb{R}^2$.*

In this section, we will examine some special examples of linear transformations in $\mathbb{R}^2$ including rotations and reflections. We will use the geometric descriptions of vector addition and scalar multiplication discussed earlier to show that a rotation of vectors through an angle and reflection of a vector across a line are examples of linear transformations.

More generally, denote a transformation given by a rotation by T. Why is such a transformation linear? Consider the following picture which illustrates a rotation. Let $\vec{u}, \vec{v}$ denote vectors.



Let's consider how to obtain $T(\vec{u} + \vec{v})$. Simply, you add $T(\vec{u})$ and $T(\vec{v})$. Here is why. If you add $T(\vec{u})$ to $T(\vec{v})$ you get the diagonal of the parallelogram determined by $T(\vec{u})$ and $T(\vec{v})$, as this action is our usual vector addition. Now, suppose we first add $\vec{u}$ and $\vec{v}$, and then apply the transformation T to $\vec{u}+\vec{v}$. Hence, we find $T(\vec{u}+\vec{v})$. As shown in the diagram, this will result in the same vector. In other words, $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$.

This is because the rotation preserves all angles between the vectors as well as their lengths. In particular, it preserves the shape of this parallelogram. Thus both $T(\vec{u}) + T(\vec{v})$ and $T(\vec{u} + \vec{v})$ give the same vector. It follows that T distributes across addition of the vectors of $\mathbb{R}^2$.

Similarly, if k is a scalar, it follows that $T(k\vec{u}) = kT(\vec{u})$. Thus rotations are an example of a linear transformation by Definition 5.2.

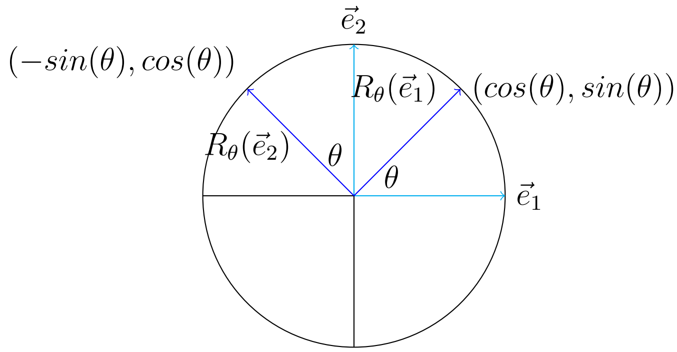
The following theorem gives the matrix of a linear transformation which rotates all vectors through an angle of θ.

Theorem 5.22: Rotation

Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation given by rotating vectors through an angle of θ. Then the matrix A of R_θ is given by

$$\left[\begin{array}{rr} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{array}\right]$$

Proof. Let $\vec{e}_1 = \left[\begin{array}{c} 1 \\ 0 \end{array}\right]$ and $\vec{e}_2 = \left[\begin{array}{c} 0 \\ 1 \end{array}\right]$. These identify the geometric vectors which point along the positive x axis and positive y axis as shown.



From Theorem 5.6, we need to find $R_\theta(\vec{e}_1)$ and $R_\theta(\vec{e}_2)$, and use these as the columns of the matrix A of T. We can use $\cos, \sin$ of the angle θ to find the coordinates of $R_\theta(\vec{e}_1)$ as shown in the above picture. The coordinates of $R_\theta(\vec{e}_2)$ also follow from trigonometry. Thus

$$R_\theta(\vec{e}_1) = \left[\begin{array}{r} \cos\theta \\ \sin\theta \end{array} \right], R_\theta(\vec{e}_2) = \left[\begin{array}{r} -\sin\theta \\ \cos\theta \end{array} \right]$$

Therefore, from Theorem 5.6,

$$A = \left[\begin{array}{rr} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array} \right]$$

We can also prove this algebraically without the use of the above picture. The definition of $(\cos(\theta), \sin(\theta))$ is as the coordinates of the point of $R_\theta(\vec{e}_1)$. Now the point of the vector $\vec{e}_2$ is exactly $\pi/2$ further along the unit circle from the point of $\vec{e}_1$, and therefore after rotation through an angle of θ the coordinates x and y of the point of $R_\theta(\vec{e}_2)$ are given by

$$(x, y) = (\cos(\theta + \pi/2), \sin(\theta + \pi/2)) = (-\sin\theta, \cos\theta)$$

♢

Consider the following example.

Example 5.23: Rotation in $\mathbb{R}^2$

Let $R_{\frac{\pi}{2}} : \mathbb{R}^2 \to \mathbb{R}^2$ denote rotation through $\pi/2$. Find the matrix of $R_{\frac{\pi}{2}}$. Then, find $R_{\frac{\pi}{2}}(\vec{x})$ where $\vec{x} = \left[\begin{array}{r} 1 \\ -2 \end{array} \right]$.

Solution. By Theorem 5.22, the matrix of $R_{\frac{\pi}{2}}$ is given by

$$\left[\begin{array}{rr} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{array} \right] = \left[\begin{array}{rr} \cos(\pi/2) & -\sin(\pi/2) \\ \sin(\pi/2) & \cos(\pi/2) \end{array} \right] = \left[\begin{array}{rr} 0 & -1 \\ 1 & 0 \end{array} \right]$$

To find $R_{\frac{\pi}{2}}(\vec{x})$, we multiply the matrix of $R_{\frac{\pi}{2}}$ by $\vec{x}$ as follows

$$\left[\begin{array}{rr} 0 & -1 \\ 1 & 0 \end{array} \right] \left[\begin{array}{r} 1 \\ -2 \end{array} \right] = \left[\begin{array}{r} 2 \\ 1 \end{array} \right]$$

□

We now look at an example of a linear transformation involving two angles.

Example 5.24: The Rotation Matrix of the Sum of Two Angles

Find the matrix of the linear transformation which is obtained by first rotating all vectors through an angle of ϕ and then through an angle θ. Hence the linear transformation rotates all vectors through an angle of $\theta + \phi$.

Solution. Let $R_{\theta+\phi}$ denote the linear transformation which rotates every vector through an angle of $\theta + \phi$. Then to obtain $R_{\theta+\phi}$, we first apply R_ϕ and then R_θ where R_ϕ is the linear transformation which rotates through an angle of ϕ and R_θ is the linear transformation which rotates through an angle of θ. Denoting the corresponding matrices by $A_{\theta+\phi}$, A_ϕ, and A_θ, it follows that for every $\vec{u}$

$$R_{\theta+\phi}(\vec{u}) = A_{\theta+\phi}\vec{u} = A_\theta A_\phi \vec{u} = R_\theta R_\phi(\vec{u})$$

Notice the order of the matrices here!

Consequently, you must have

$$\begin{aligned} A_{\theta+\phi} &= \left[\begin{array}{cc} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{array}\right] \\ &= \left[\begin{array}{cc} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array}\right] \left[\begin{array}{cc} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{array}\right] = A_\theta A_\phi \end{aligned}$$

The usual matrix multiplication yields

$$\begin{aligned} A_{\theta+\phi} &= \left[\begin{array}{cc} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{array}\right] \\ &= \left[\begin{array}{cc} \cos\theta\cos\phi - \sin\theta\sin\phi & -\cos\theta\sin\phi - \sin\theta\cos\phi \\ \sin\theta\cos\phi + \cos\theta\sin\phi & \cos\theta\cos\phi - \sin\theta\sin\phi \end{array}\right] \\ &= A_\theta A_\phi \end{aligned}$$

Don't these look familiar? They are the usual trigonometric identities for the sum of two angles derived here using linear algebra concepts.

□

Here we have focused on rotations in two dimensions. However, you can consider rotations and other geometric concepts in any number of dimensions. This is one of the major advantages of linear algebra. You can break down a difficult geometrical procedure into small steps, each corresponding to multiplication by an appropriate matrix. Then by multiplying the matrices, you can obtain a single matrix which can give you numerical information on the results of applying the given sequence of simple procedures.

Linear transformations which reflect vectors across a line are a second important type of transformations in $\mathbb{R}^2$. Consider the following theorem.

Theorem 5.25: Reflection

Let $Q_m : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation given by reflecting vectors over the line $\vec{y} = m\vec{x}$. Then the matrix of Q_m is given by

$$\frac{1}{1+m^2}\left[\begin{array}{cc} 1-m^2 & 2m \\ 2m & m^2-1 \end{array}\right]$$

Consider the following example.

Example 5.26: Reflection in $\mathbb{R}^2$

Let $Q_2 : \mathbb{R}^2 \to \mathbb{R}^2$ denote reflection over the line $\vec{y} = 2\vec{x}$. Then Q_2 is a linear transformation. Find the matrix of Q_2. Then, find $Q_2(\vec{x})$ where $\vec{x} = \left[\begin{array}{c} 1 \\ -2 \end{array}\right]$.

Solution. By Theorem 5.25, the matrix of Q_2 is given by

$$\frac{1}{1+m^2}\left[\begin{array}{cc} 1-m^2 & 2m \\ 2m & m^2-1 \end{array}\right] = \frac{1}{1+(2)^2}\left[\begin{array}{cc} 1-(2)^2 & 2(2) \\ 2(2) & (2)^2-1 \end{array}\right] = \frac{1}{5}\left[\begin{array}{rr} -3 & 8 \\ 8 & 3 \end{array}\right]$$

To find $Q_2(\vec{x})$ we multiply $\vec{x}$ by the matrix of Q_2 as follows:

$$\frac{1}{5}\left[\begin{array}{rr} -3 & 8 \\ 8 & 3 \end{array}\right]\left[\begin{array}{r} 1 \\ -2 \end{array}\right] = \left[\begin{array}{r} -\frac{19}{5} \\ \frac{2}{5} \end{array}\right]$$

□

Consider the following example which incorporates a reflection as well as a rotation of vectors.

Example 5.27: Rotation Followed by a Reflection

Find the matrix of the linear transformation which is obtained by first rotating all vectors through an angle of $\pi/6$ and then reflecting through the x axis.

Solution. By Theorem 5.22, the matrix of the transformation which involves rotating through an angle of $\pi/6$ is

$$\left[\begin{array}{cc} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{array}\right] = \left[\begin{array}{cc} \frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}\sqrt{3} \end{array}\right]$$

Reflecting across the x axis is the same action as reflecting vectors over the line $\vec{y} = m\vec{x}$ with $m = 0$. By Theorem 5.25, the matrix for the transformation which reflects all vectors through the x axis is

$$\frac{1}{1+m^2}\left[\begin{array}{cc} 1-m^2 & 2m \\ 2m & m^2-1 \end{array}\right] = \frac{1}{1+(0)^2}\left[\begin{array}{cc} 1-(0)^2 & 2(0) \\ 2(0) & (0)^2-1 \end{array}\right] = \left[\begin{array}{rr} 1 & 0 \\ 0 & -1 \end{array}\right]$$

Therefore, the matrix of the linear transformation which first rotates through $\pi/6$ and then reflects through the x axis is given by

$$\left[\begin{array}{rr} 1 & 0 \\ 0 & -1 \end{array}\right]\left[\begin{array}{rr} \frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}\sqrt{3} \end{array}\right]=\left[\begin{array}{rr} \frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \end{array}\right]$$

♦

Exercises

Exercise 5.4.1 *Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/3$.*

Exercise 5.4.2 *Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/4$.*

Exercise 5.4.3 *Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $-\pi/3$.*

Exercise 5.4.4 *Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $2\pi/3$.*

Exercise 5.4.5 *Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/12$.* ***Hint:*** *Note that $\pi/12 = \pi/3 - \pi/4$.*

Exercise 5.4.6 *Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $2\pi/3$ and then reflects across the x axis.*

Exercise 5.4.7 *Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/3$ and then reflects across the x axis.*

Exercise 5.4.8 *Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/4$ and then reflects across the x axis.*

Exercise 5.4.9 *Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/6$ and then reflects across the x axis followed by a reflection across the y axis.*

Exercise 5.4.10 *Find the matrix for the linear transformation which reflects every vector in $\mathbb{R}^2$ across the x axis and then rotates every vector through an angle of $\pi/4$.*

Exercise 5.4.11 *Find the matrix for the linear transformation which reflects every vector in $\mathbb{R}^2$ across the y axis and then rotates every vector through an angle of $\pi/4$.*

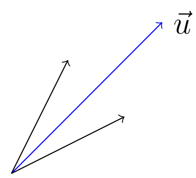
Exercise 5.4.12 *Find the matrix for the linear transformation which reflects every vector in* $\mathbb{R}^2$ *across the* x *axis and then rotates every vector through an angle of* $\pi/6$.

Exercise 5.4.13 *Find the matrix for the linear transformation which reflects every vector in* $\mathbb{R}^2$ *across the* y *axis and then rotates every vector through an angle of* $\pi/6$.

Exercise 5.4.14 *Find the matrix for the linear transformation which rotates every vector in* $\mathbb{R}^2$ *through an angle of* $5\pi/12$. ***Hint:*** *Note that* $5\pi/12 = 2\pi/3 - \pi/4$.

Exercise 5.4.15 *Find the matrix of the linear transformation which rotates every vector in* $\mathbb{R}^3$ *counter clockwise about the* z *axis when viewed from the positive* z *axis through an angle of 30° and then reflects through the* xy *plane.*

Exercise 5.4.16 *Let* $\vec{u} = \left[\begin{array}{c} a \\ b \end{array}\right]$ *be a unit vector in* $\mathbb{R}^2$. *Find the matrix which reflects all vectors across this vector, as shown in the following picture.*



Hint: *Notice that* $\left[\begin{array}{c} a \\ b \end{array}\right] = \left[\begin{array}{c} \cos\theta \\ \sin\theta \end{array}\right]$ *for some* θ. *First rotate through* $-\theta$. *Next reflect through the* x *axis. Finally rotate through* θ.

5.5 One to One and Onto Transformations

Outcomes

A. *Determine if a linear transformation is onto or one to one.*

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear transformation. We define the **range** or **image** of T as the set of vectors of $\mathbb{R}^m$ which are of the form $T(\vec{x})$ (equivalently, $A\vec{x}$) for some $\vec{x} \in \mathbb{R}^n$. It is common to write $T\mathbb{R}^n$, $T(\mathbb{R}^n)$, or $\mathrm{Im}(T)$ to denote these vectors.

Lemma 5.28: Range of a Matrix Transformation

Let A be an $m \times n$ matrix where $A_1, \cdots, A_n$ denote the columns of A. Then, for a vector $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ in $\mathbb{R}^n$,

$$A\vec{x} = \sum_{k=1}^{n} x_k A_k$$

Therefore, $A(\mathbb{R}^n)$ is the collection of all linear combinations of these products.

Proof. This follows from the definition of matrix multiplication. □

This section is devoted to studying two important characterizations of linear transformations, called one to one and onto. We define them now.

Definition 5.29: One to One

Suppose $\vec{x}_1$ and $\vec{x}_2$ are vectors in $\mathbb{R}^n$. A linear transformation $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ is called ***one to one*** *(often written as $1-1$) if whenever $\vec{x}_1 \neq \vec{x}_2$ it follows that :*

$$T(\vec{x}_1) \neq T(\vec{x}_2)$$

Equivalently, if $T(\vec{x}_1) = T(\vec{x}_2)$, then $\vec{x}_1 = \vec{x}_2$. Thus, T is one to one if it never takes two different vectors to the same vector.

The second important characterization is called onto.

Definition 5.30: Onto

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear transformation. Then T is called ***onto*** *if whenever $\vec{x}_2 \in \mathbb{R}^m$ there exists $\vec{x}_1 \in \mathbb{R}^n$ such that $T(\vec{x}_1) = \vec{x}_2$.*

We often call a linear transformation which is one-to-one an **injection**. Similarly, a linear transformation which is onto is often called a **surjection**.

The following proposition is an important result.

Proposition 5.31: One to One

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear transformation. Then T is one to one if and only if $T(\vec{x}) = \vec{0}$ implies $\vec{x} = \vec{0}$.

Proof. We need to prove two things here. First, we will prove that if T is one to one, then $T(\vec{x}) = \vec{0}$ implies that $\vec{x} = \vec{0}$. Second, we will show that if $T(\vec{x}) = \vec{0}$ implies that $\vec{x} = \vec{0}$, then it follows that T is one to one. Recall that a linear transformation has the property that $T(\vec{0}) = \vec{0}$.

Suppose first that T is one to one and consider $T(\vec{0})$.

$$T(\vec{0}) = T\left(\vec{0} + \vec{0}\right) = T(\vec{0}) + T(\vec{0})$$

and so, adding the additive inverse of $T(\vec{0})$ to both sides, one sees that $T(\vec{0}) = \vec{0}$. If $T(\vec{x}) = \vec{0}$ it must be the case that $\vec{x} = \vec{0}$ because it was just shown that $T(\vec{0}) = \vec{0}$ and T is assumed to be one to one.

Now assume that if $T(\vec{x}) = \vec{0}$, then it follows that $\vec{x} = \vec{0}$. If $T(\vec{v}) = T(\vec{u})$, then

$$T(\vec{v}) - T(\vec{u}) = T(\vec{v} - \vec{u}) = \vec{0}$$

which shows that $\vec{v} - \vec{u} = 0$. In other words, $\vec{v} = \vec{u}$, and T is one to one. $\square$

Note that this proposition says that if $A = \left[\begin{array}{ccc} A_1 & \cdots & A_n \end{array}\right]$ then A is one to one if and only if whenever

$$0 = \sum_{k=1}^{n} c_k A_k$$

it follows that each scalar $c_k = 0$.

We will now take a look at an example of a one to one and onto linear transformation.

Example 5.32: A One to One and Onto Linear Transformation

Suppose

$$T\left[\begin{array}{c} x \\ y \end{array}\right] = \left[\begin{array}{cc} 1 & 1 \\ 1 & 2 \end{array}\right]\left[\begin{array}{c} x \\ y \end{array}\right]$$

Then, $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation. Is T onto? Is it one to one?

Solution. Recall that because T can be expressed as matrix multiplication, we know that T is a linear transformation. We will start by looking at onto. So suppose $\left[\begin{array}{c} a \\ b \end{array}\right] \in \mathbb{R}^2$. Does there exist $\left[\begin{array}{c} x \\ y \end{array}\right] \in \mathbb{R}^2$ such that $T\left[\begin{array}{c} x \\ y \end{array}\right] = \left[\begin{array}{c} a \\ b \end{array}\right]$? If so, then since $\left[\begin{array}{c} a \\ b \end{array}\right]$ is an arbitrary vector in $\mathbb{R}^2$, it will follow that T is onto.

This question is familiar to you. It is asking whether there is a solution to the equation

$$\left[\begin{array}{cc} 1 & 1 \\ 1 & 2 \end{array}\right]\left[\begin{array}{c} x \\ y \end{array}\right] = \left[\begin{array}{c} a \\ b \end{array}\right]$$

This is the same thing as asking for a solution to the following system of equations.

$$\begin{array}{c} x + y = a \\ x + 2y = b \end{array}$$

Set up the augmented matrix and row reduce.

$$\left[\begin{array}{cc|c} 1 & 1 & a \\ 1 & 2 & b \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 2a - b \\ 0 & 1 & b - a \end{array}\right] \tag{5.4}$$

You can see from this point that the system has a solution. Therefore, we have shown that for any a, b, there is a $\left[\begin{array}{c} x \\ y \end{array}\right]$ such that $T\left[\begin{array}{c} x \\ y \end{array}\right]=\left[\begin{array}{c} a \\ b \end{array}\right]$. Thus T is onto.

Now we want to know if T is one to one. By Proposition 5.31 it is enough to show that $A\vec{x}=0$ implies $\vec{x}=0$. Consider the system $A\vec{x}=0$ given by:

$$\left[\begin{array}{cc} 1 & 1 \\ 1 & 2 \end{array}\right]\left[\begin{array}{c} x \\ y \end{array}\right]=\left[\begin{array}{c} 0 \\ 0 \end{array}\right]$$

This is the same as the system given by

$$\begin{array}{c} x+y=0 \\ x+2y=0 \end{array}$$

We need to show that the solution to this system is $x=0$ and $y=0$. By setting up the augmented matrix and row reducing, we end up with

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

This tells us that $x=0$ and $y=0$. Returning to the original system, this says that if

$$\left[\begin{array}{cc} 1 & 1 \\ 1 & 2 \end{array}\right]\left[\begin{array}{c} x \\ y \end{array}\right]=\left[\begin{array}{c} 0 \\ 0 \end{array}\right]$$

then

$$\left[\begin{array}{c} x \\ y \end{array}\right]=\left[\begin{array}{c} 0 \\ 0 \end{array}\right]$$

In other words, $A\vec{x}=0$ implies that $\vec{x}=0$. By Proposition 5.31, A is one to one, and so T is also one to one.

We also could have seen that T is one to one from our above solution for onto. By looking at the matrix given by 5.4, you can see that there is a **unique** solution given by $x=2a-b$ and $y=b-a$. Therefore, there is only one vector, specifically $\left[\begin{array}{c} x \\ y \end{array}\right]=\left[\begin{array}{c} 2a-b \\ b-a \end{array}\right]$ such that $T\left[\begin{array}{c} x \\ y \end{array}\right]=\left[\begin{array}{c} a \\ b \end{array}\right]$. Hence by Definition 5.29, T is one to one. ∎

Example 5.33: An Onto Transformation

Let $T:\mathbb{R}^4 \mapsto \mathbb{R}^2$ be a linear transformation defined by

$$T\left[\begin{array}{c} a \\ b \\ c \\ d \end{array}\right]=\left[\begin{array}{c} a+d \\ b+c \end{array}\right] \text{ for all } \left[\begin{array}{c} a \\ b \\ c \\ d \end{array}\right] \in \mathbb{R}^4$$

Prove that T is onto but not one to one.

Solution. You can prove that T is in fact linear.

To show that T is onto, let $\begin{bmatrix} x \\ y \end{bmatrix}$ be an arbitrary vector in $\mathbb{R}^2$. Taking the vector $\begin{bmatrix} x \\ y \\ 0 \\ 0 \end{bmatrix} \in \mathbb{R}^4$ we have

$$T \begin{bmatrix} x \\ y \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} x+0 \\ y+0 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

This shows that T is onto.

By Proposition 5.31 T is one to one if and only if $T(\vec{x}) = \vec{0}$ implies that $\vec{x} = \vec{0}$. Observe that

$$T \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1+-1 \\ 0+0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

There exists a nonzero vector $\vec{x}$ in $\mathbb{R}^4$ such that $T(\vec{x}) = \vec{0}$. It follows that T is not one to one. □

The above examples demonstrate a method to determine if a linear transformation T is one to one or onto. It turns out that the matrix A of T can provide this information.

Theorem 5.34: Matrix of a One to One or Onto Transformation

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear transformation induced by the $m \times n$ matrix A. Then T is one to one if and only if the rank of A is n. T is onto if and only if the rank of A is m.

Consider Example 5.33. Above we showed that T was onto but not one to one. We can now use this theorem to determine this fact about T.

Example 5.35: An Onto Transformation

Let $T : \mathbb{R}^4 \mapsto \mathbb{R}^2$ be a linear transformation defined by

$$T \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} a+d \\ b+c \end{bmatrix} \text{ for all } \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \in \mathbb{R}^4$$

Prove that T is onto but not one to one.

Solution. Using Theorem 5.34 we can show that T is onto but not one to one from the matrix of T. Recall that to find the matrix A of T, we apply T to each of the standard basis vectors $\vec{e}_i$ of $\mathbb{R}^4$. The result is the 2×4 matrix A given by

$$A = \left[\begin{array}{cccc} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right]$$

Fortunately, this matrix is already in reduced row-echelon form. The rank of A is 2. Therefore by the above theorem T is onto but not one to one. $\square$

Recall that if S and T are linear transformations, we can discuss their composite denoted $S \circ T$. The following examines what happens if both S and T are onto.

Example 5.36: Composite of Onto Transformations

Let $T : \mathbb{R}^k \mapsto \mathbb{R}^n$ and $S : \mathbb{R}^n \mapsto \mathbb{R}^m$ be linear transformations. If T and S are onto, then $S \circ T$ is onto.

Solution. Let $\vec{z} \in \mathbb{R}^m$. Since S is onto, there exists a vector $\vec{y} \in \mathbb{R}^n$ such that $S(\vec{y}) = \vec{z}$. Furthermore, since T is onto, there exists a vector $\vec{x} \in \mathbb{R}^k$ such that $T(\vec{x}) = \vec{y}$. Thus

$$\vec{z} = S(\vec{y}) = S(T(\vec{x})) = (ST)(\vec{x}),$$

showing that for each $\vec{z} \in \mathbb{R}^m$ there exists and $\vec{x} \in \mathbb{R}^k$ such that $(ST)(\vec{x}) = \vec{z}$. Therefore, $S \circ T$ is onto. $\square$

The next example shows the same concept with regards to one-to-one transformations.

Example 5.37: Composite of One to One Transformations

Let $T : \mathbb{R}^k \mapsto \mathbb{R}^n$ and $S : \mathbb{R}^n \mapsto \mathbb{R}^m$ be linear transformations. Prove that if T and S are one to one, then $S \circ T$ is one-to-one.

Solution. To prove that $S \circ T$ is one to one, we need to show that if $S(T(\vec{v})) = \vec{0}$ it follows that $\vec{v} = \vec{0}$. Suppose that $S(T(\vec{v})) = \vec{0}$. Since S is one to one, it follows that $T(\vec{v}) = \vec{0}$. Similarly, since T is one to one, it follows that $\vec{v} = \vec{0}$. Hence $S \circ T$ is one to one.

$\square$

Exercises

Exercise 5.5.1 *Let T be a linear transformation given by*

$$T \left[\begin{array}{c} x \\ y \end{array}\right] = \left[\begin{array}{cc} 2 & 1 \\ 0 & 1 \end{array}\right] \left[\begin{array}{c} x \\ y \end{array}\right]$$

Is T one to one? Is T onto?

Exercise 5.5.2 *Let T be a linear transformation given by*

$$T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ 2 & 1 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Is T one to one? Is T onto?

Exercise 5.5.3 *Let T be a linear transformation given by*

$$T \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Is T one to one? Is T onto?

Exercise 5.5.4 *Let T be a linear transformation given by*

$$T \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 3 & -5 \\ 2 & 0 & 2 \\ 2 & 4 & -6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Is T one to one? Is T onto?

Exercise 5.5.5 *Give an example of a 3×2 matrix with the property that the linear transformation determined by this matrix is one to one but not onto.*

Exercise 5.5.6 *Suppose A is an $m \times n$ matrix in which $m \leq n$. Suppose also that the rank of A equals m. Show that the transformation T determined by A maps $\mathbb{R}^n$ onto $\mathbb{R}^m$.* ***Hint:*** *The vectors $\vec{e}_1, \cdots, \vec{e}_m$ occur as columns in the reduced row-echelon form for A.*

Exercise 5.5.7 *Suppose A is an $m \times n$ matrix in which $m \geq n$. Suppose also that the rank of A equals n. Show that A is one to one.* ***Hint:*** *If not, there exists a vector, $\vec{x}$ such that $A\vec{x} = 0$, and this implies at least one column of A is a linear combination of the others. Show this would require the rank to be less than n.*

Exercise 5.5.8 *Explain why an $n \times n$ matrix A is both one to one and onto if and only if its rank is n.*

5.6 The General Solution of a Linear System

Outcomes

A. *Use linear transformations to determine the particular solution and general solution to a system of equations.*

B. *Find the kernel of a linear transformation.*

Recall the definition of a linear transformation discussed above. T is a **linear transformation** if whenever $\vec{x}, \vec{y}$ are vectors and k, p are scalars,

$$T\left(k\vec{x} + p\vec{y}\right) = kT\left(\vec{x}\right) + pT\left(\vec{y}\right)$$

Thus linear transformations distribute across addition and pass scalars to the outside.

It turns out that we can use linear transformations to solve linear systems of equations. Indeed given a system of linear equations of the form $A\vec{x} = \vec{b}$, one may rephrase this as $T(\vec{x}) = \vec{b}$ where T is the linear transformation T_A induced by the coefficient matrix A. With this in mind consider the following definition.

Definition 5.38: Particular Solution of a System of Equations

Suppose a linear system of equations can be written in the form

$$T\left(\vec{x}\right) = \vec{b}$$

If $T\left(\vec{x}_p\right) = \vec{b}$, then $\vec{x}_p$ is called a ***particular solution*** *of the linear system.*

Recall that a system is called homogeneous if every equation in the system is equal to 0. Suppose we represent a homogeneous system of equations by $T\left(\vec{x}\right) = 0$. It turns out that the $\vec{x}$ for which $T\left(\vec{x}\right) = 0$ are part of a special set called the **null space** of T. We may also refer to the null space as the **kernel** of T, and we write $ker\left(T\right)$.

Consider the following definition.

Definition 5.39: Null Space or Kernel of a Linear Transformation

Let T be a linear transformation. Define

$$\ker\left(T\right) = \left\{\vec{x} : T\left(\vec{x}\right) = \vec{0}\right\}$$

The kernel, $\ker\left(T\right)$ *consists of the set of all vectors $\vec{x}$ for which $T(\vec{x}) = \vec{0}$. This is also called the* ***null space*** *of T.*

We may also refer to the kernel of T as the **solution space** of the equation $T\left(\vec{x}\right) = \vec{0}$.

Consider the following example.

Example 5.40: The Kernel of the Derivative

Let $\frac{d}{dx}$ denote the linear transformation defined on f, the functions which are defined on $\mathbb{R}$ and have a continuous derivative. Find $\ker\left(\frac{d}{dx}\right)$.

Solution. The example asks for functions f which the property that $\frac{df}{dx} = 0$. As you may know from calculus, these functions are the constant functions. Thus $\ker\left(\frac{d}{dx}\right)$ is the set of constant functions. ☐

Definition 5.39 states that $\ker(T)$ is the set of solutions to the equation,

$$T(\vec{x}) = \vec{0}$$

Since we can write $T(\vec{x})$ as $A\vec{x}$, you have been solving such equations for quite some time.

We have spent a lot of time finding solutions to systems of equations in general, as well as homogeneous systems. Suppose we look at a system given by $A\vec{x} = \vec{b}$, and consider the related homogeneous system. By this, we mean that we replace $\vec{b}$ by $\vec{0}$ and look at $A\vec{x} = \vec{0}$. It turns out that there is a very important relationship between the solutions of the original system and the solutions of the associated homogeneous system. In the following theorem, we use linear transformations to denote a system of equations. Remember that $T(\vec{x}) = A\vec{x}$.

Theorem 5.41: Particular Solution and General Solution

Suppose $\vec{x}_p$ is a solution to the linear system given by ,

$$T(\vec{x}) = \vec{b}$$

Then if $\vec{y}$ is any other solution to $T(\vec{x}) = \vec{b}$, there exists $\vec{x}_0 \in \ker(T)$ such that

$$\vec{y} = \vec{x}_p + \vec{x}_0$$

Hence, every solution to the linear system can be written as a sum of a particular solution, $\vec{x}_p$, and a solution $\vec{x}_0$ to the associated homogeneous system given by $T(\vec{x}) = \vec{0}$.

Proof. Consider $\vec{y} - \vec{x}_p = \vec{y} + (-1)\vec{x}_p$. Then $T(\vec{y} - \vec{x}_p) = T(\vec{y}) - T(\vec{x}_p)$. Since $\vec{y}$ and $\vec{x}_p$ are both solutions to the system, it follows that $T(\vec{y}) = \vec{b}$ and $T(\vec{x}_p) = \vec{b}$.

Hence, $T(\vec{y}) - T(\vec{x}_p) = \vec{b} - \vec{b} = \vec{0}$. Let $\vec{x}_0 = \vec{y} - \vec{x}_p$. Then, $T(\vec{x}_0) = \vec{0}$ so $\vec{x}_0$ is a solution to the associated homogeneous system and so is in $\ker(T)$. ∎

Sometimes people remember the above theorem in the following form. The solutions to the system $T(\vec{x}) = \vec{b}$ are given by $\vec{x}_p + \ker(T)$ where $\vec{x}_p$ is a particular solution to $T(\vec{x}) = \vec{b}$.

For now, we have been speaking about the kernel or null space of a linear transformation T. However, we know that every linear transformation T is determined by some matrix A. Therefore, we can also speak about the null space of a matrix. Consider the following example.

Example 5.42: The Null Space of a Matrix

Let

$$A = \left[\begin{array}{cccc} 1 & 2 & 3 & 0 \\ 2 & 1 & 1 & 2 \\ 4 & 5 & 7 & 2 \end{array}\right]$$

Find $\mathrm{null}(A)$. Equivalently, find the solutions to the system of equations $A\vec{x} = \vec{0}$.

Solution. We are asked to find $\left\{\vec{x} : A\vec{x} = \vec{0}\right\}$. In other words we want to solve the system, $A\vec{x} = \vec{0}$. Let $\vec{x} = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$. Then this amounts to solving

$$\begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & 1 & 1 & 2 \\ 4 & 5 & 7 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This is the linear system

$$\begin{gathered} x + 2y + 3z = 0 \\ 2x + y + z + 2w = 0 \\ 4x + 5y + 7z + 2w = 0 \end{gathered}$$

To solve, set up the augmented matrix and row reduce to find the reduced row-echelon form.

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & 0 & 0 \\ 2 & 1 & 1 & 2 & 0 \\ 4 & 5 & 7 & 2 & 0 \end{array}\right] \rightarrow \cdots \rightarrow \left[\begin{array}{cccc|c} 1 & 0 & -\frac{1}{3} & \frac{4}{3} & 0 \\ 0 & 1 & \frac{5}{3} & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

This yields $x = \frac{1}{3}z - \frac{4}{3}w$ and $y = \frac{2}{3}w - \frac{5}{3}z$. Since $\operatorname{null}(A)$ consists of the solutions to this system, it consists vectors of the form,

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

$\square$

Consider the following example.

Example 5.43: A General Solution

The ***general solution*** *of a linear system of equations is the set of all possible solutions. Find the general solution to the linear system,*

$$\begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & 1 & 1 & 2 \\ 4 & 5 & 7 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 9 \\ 7 \\ 25 \end{bmatrix}$$

given that $\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}$ *is one solution.*

Solution. Note the matrix of this system is the same as the matrix in Example 5.42. Therefore, from Theorem 5.41, you will obtain all solutions to the above linear system by adding a particular solution $\vec{x}_p$ to the solutions of the associated homogeneous system, $\vec{x}$. One particular solution is given above by

$$\vec{x}_p = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix} \tag{5.5}$$

Using this particular solution along with the solutions found in Example 5.42, we obtain the following solutions,

$$z \begin{bmatrix} \frac{1}{3} \\ -\frac{5}{3} \\ 1 \\ 0 \end{bmatrix} + w \begin{bmatrix} -\frac{4}{3} \\ \frac{2}{3} \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}$$

Hence, any solution to the above linear system is of this form. ◻

Exercises

Exercise 5.6.1 *Write the solution set of the following system as a linear combination of vectors*

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

Exercise 5.6.2 *Using Problem 5.6.1 find the general solution to the following linear system.*

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

Exercise 5.6.3 *Write the solution set of the following system as a linear combination of vectors*

$$\begin{bmatrix} 0 & -1 & 2 \\ 1 & -2 & 1 \\ 1 & -4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Exercise 5.6.4 *Using Problem 5.6.3 find the general solution to the following linear system.*

$$\begin{bmatrix} 0 & -1 & 2 \\ 1 & -2 & 1 \\ 1 & -4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

Exercise 5.6.5 *Write the solution set of the following system as a linear combination of vectors.*

$$\begin{bmatrix} 1 & -1 & 2 \\ 1 & -2 & 0 \\ 3 & -4 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Exercise 5.6.6 *Using Problem 5.6.5 find the general solution to the following linear system.*

$$\begin{bmatrix} 1 & -1 & 2 \\ 1 & -2 & 0 \\ 3 & -4 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

Exercise 5.6.7 *Write the solution set of the following system as a linear combination of vectors*

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

Exercise 5.6.8 *Using Problem 5.6.7 find the general solution to the following linear system.*

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

Exercise 5.6.9 *Write the solution set of the following system as a linear combination of vectors*

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

Exercise 5.6.10 *Using Problem 5.6.9 find the general solution to the following linear system.*

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & -1 & 1 & 0 \\ 3 & -1 & 3 & 2 \\ 3 & 3 & 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 3 \end{bmatrix}$$

Exercise 5.6.11 *Write the solution set of the following system as a linear combination of vectors*

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

Exercise 5.6.12 *Using Problem 5.6.11 find the general solution to the following linear system.*

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

Exercise 5.6.13 *Write the solution set of the following system as a linear combination of vectors*

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

Exercise 5.6.14 *Using Problem 5.6.13 find the general solution to the following linear system.*

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & -1 & 1 & 0 \\ 3 & 1 & 1 & 2 \\ 3 & 3 & 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 3 \end{bmatrix}$$

Exercise 5.6.15 *Write the solution set of the following system as a linear combination of vectors*

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 2 & 1 & 1 & 2 \\ 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Exercise 5.6.16 *Using Problem 5.6.15 find the general solution to the following linear system.*

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

Exercise 5.6.17 *Suppose $A\vec{x} = \vec{b}$ has a solution. Explain why the solution is unique precisely when $A\vec{x} = \vec{0}$ has only the trivial solution.*

6. Spectral Theory

6.1 Eigenvalues and Eigenvectors of a Matrix

Outcomes

A. *Describe eigenvalues geometrically and algebraically.*

B. *Find eigenvalues and eigenvectors for a square matrix.*

Spectral Theory refers to the study of eigenvalues and eigenvectors of a matrix. It is of fundamental importance in many areas and is the subject of our study for this chapter.

6.1.1. Definition of Eigenvectors and Eigenvalues

In this section, we will work with the entire set of complex numbers, denoted by $\mathbb{C}$. Recall that the real numbers, $\mathbb{R}$ are contained in the complex numbers, so the discussions in this section apply to both real and complex numbers.

To illustrate the idea behind what will be discussed, consider the following example.

Example 6.1: Eigenvectors and Eigenvalues

Let

$$A = \left[\begin{array}{rrr} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{array}\right]$$

Compute the product AX for

$$X = \left[\begin{array}{r} 5 \\ -4 \\ 3 \end{array}\right], X = \left[\begin{array}{r} 1 \\ 0 \\ 0 \end{array}\right]$$

What do you notice about AX in each of these products?

Solution. First, compute AX for

$$X = \left[\begin{array}{r} 5 \\ -4 \\ 3 \end{array}\right]$$

This product is given by

$$AX = \begin{bmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{bmatrix} \begin{bmatrix} -5 \\ -4 \\ 3 \end{bmatrix} = \begin{bmatrix} -50 \\ -40 \\ 30 \end{bmatrix} = 10 \begin{bmatrix} -5 \\ -4 \\ 3 \end{bmatrix}$$

In this case, the product AX resulted in a vector which is equal to 10 times the vector X. In other words, $AX = 10X$.

Let's see what happens in the next product. Compute AX for the vector

$$X = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

This product is given by

$$AX = \begin{bmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

In this case, the product AX resulted in a vector equal to 0 times the vector X, $AX = 0X$.

Perhaps this matrix is such that AX results in kX, for every vector X. However, consider

$$\begin{bmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -5 \\ 38 \\ -11 \end{bmatrix}$$

In this case, AX did not result in a vector of the form kX for some scalar k. ◻

There is something special about the first two products calculated in Example 6.1. Notice that for each, $AX = kX$ where k is some scalar. When this equation holds for some X and k, we call the scalar k an **eigenvalue** of A. We often use the special symbol λ instead of k when referring to eigenvalues. In Example 6.1, the values 10 and 0 are eigenvalues for the matrix A and we can label these as $\lambda_1 = 10$ and $\lambda_2 = 0$.

When $AX = \lambda X$ for some $X \neq 0$, we call such an X an **eigenvector** of the matrix A. The eigenvectors of A are associated to an eigenvalue. Hence, if λ_1 is an eigenvalue of A and $AX = \lambda_1 X$, we can label this eigenvector as X_1. Note again that in order to be an eigenvector, X must be nonzero.

There is also a geometric significance to eigenvectors. When you have a **nonzero** vector which, when multiplied by a matrix results in another vector which is parallel to the first or equal to $\mathbf{0}$, this vector is called an eigenvector of the matrix. This is the meaning when the vectors are in $\mathbb{R}^n$.

The formal definition of eigenvalues and eigenvectors is as follows.

Definition 6.2: Eigenvalues and Eigenvectors

Let A be an $n \times n$ matrix and let $X \in \mathbb{C}^n$ be a ***nonzero vector*** *for which*

$$AX = \lambda X \tag{6.1}$$

for some scalar λ. Then λ is called an ***eigenvalue*** *of the matrix A and X is called an* ***eigenvector*** *of A associated with λ, or a λ-eigenvector of A.*
The set of all eigenvalues of an $n \times n$ matrix A is denoted by $\sigma(A)$ and is referred to as the ***spectrum*** *of A.*

The eigenvectors of a matrix A are those vectors X for which multiplication by A results in a vector in the same direction or opposite direction to X. Since the zero vector 0 has no direction this would make no sense for the zero vector. As noted above, 0 is never allowed to be an eigenvector.

Let's look at eigenvectors in more detail. Suppose X satisfies 6.1. Then

$$\begin{gathered} AX - \lambda X = 0 \\ \text{or} \\ (A - \lambda I)\, X = 0 \end{gathered}$$

for some $X \neq 0$. Equivalently you could write $(\lambda I - A)\, X = 0$, which is more commonly used. Hence, when we are looking for eigenvectors, we are looking for nontrivial solutions to this homogeneous system of equations!

Recall that the solutions to a homogeneous system of equations consist of basic solutions, and the linear combinations of those basic solutions. In this context, we call the basic solutions of the equation $(\lambda I - A)\, X = 0$ **basic eigenvectors**. It follows that any (nonzero) linear combination of basic eigenvectors is again an eigenvector.

Suppose the matrix $(\lambda I - A)$ is invertible, so that $(\lambda I - A)^{-1}$ exists. Then the following equation would be true.

$$\begin{aligned} X &= IX \\ &= \left((\lambda I - A)^{-1} (\lambda I - A) \right) X \\ &= (\lambda I - A)^{-1} \left((\lambda I - A)\, X \right) \\ &= (\lambda I - A)^{-1}\, 0 \\ &= 0 \end{aligned}$$

This claims that $X = 0$. However, we have required that $X \neq 0$. Therefore $(\lambda I - A)$ cannot have an inverse!

Recall that if a matrix is not invertible, then its determinant is equal to 0. Therefore we can conclude that

$$\det(\lambda I - A) = 0 \tag{6.2}$$

Note that this is equivalent to $\det(A - \lambda I) = 0$.

The expression $\det(xI - A)$ is a polynomial (in the variable x) called the **characteristic polynomial of** A, and $\det(xI - A) = 0$ is called the **characteristic equation**. For this reason we may also refer to the eigenvalues of A as **characteristic values**, but the former is often used for historical reasons.

The following theorem claims that the roots of the characteristic polynomial are the eigenvalues of A. Thus when 6.2 holds, A has a nonzero eigenvector.

Theorem 6.3: The Existence of an Eigenvector

Let A be an $n \times n$ matrix and suppose $\det(\lambda I - A) = 0$ for some $\lambda \in \mathbb{C}$. Then λ is an eigenvalue of A and thus there exists a nonzero vector $X \in \mathbb{C}^n$ such that $AX = \lambda X$.

Proof. For A an $n \times n$ matrix, the method of Laplace Expansion demonstrates that $\det(\lambda I - A)$ is a polynomial of degree n. As such, the equation 6.2 has a solution $\lambda \in \mathbb{C}$ by the Fundamental Theorem of Algebra. The fact that λ is an eigenvalue is left as an exercise. □

6.1.2. Finding Eigenvectors and Eigenvalues

Now that eigenvalues and eigenvectors have been defined, we will study how to find them for a matrix A.

First, consider the following definition.

Definition 6.4: Multiplicity of an Eigenvalue

Let A be an $n \times n$ matrix with characteristic polynomial given by $\det(xI - A)$. Then, the multiplicity of an eigenvalue λ of A is the number of times λ occurs as a root of that characteristic polynomial.

For example, suppose the characteristic polynomial of A is given by $(x-2)^2$. Solving for the roots of this polynomial, we set $(x-2)^2 = 0$ and solve for x. We find that $\lambda = 2$ is a root that occurs twice. Hence, in this case, $\lambda = 2$ is an eigenvalue of A of multiplicity equal to 2.

We will now look at how to find the eigenvalues and eigenvectors for a matrix A in detail. The steps used are summarized in the following procedure.

Procedure 6.5: Finding Eigenvalues and Eigenvectors

Let A be an $n \times n$ matrix.

1. *First, find the eigenvalues λ of A by solving the equation $\det(xI - A) = 0$.*
2. *For each λ, find the basic eigenvectors $X \neq 0$ by finding the basic solutions to $(\lambda I - A)X = 0$.*

To verify your work, make sure that $AX = \lambda X$ for each λ and associated eigenvector X.

We will explore these steps further in the following example.

Example 6.6: Find the Eigenvalues and Eigenvectors

Let $A = \left[\begin{array}{rr} -5 & 2 \\ -7 & 4 \end{array}\right]$. *Find its eigenvalues and eigenvectors.*

Solution. We will use Procedure 6.5. First we find the eigenvalues of A by solving the equation

$$\det\left(xI - A\right) = 0$$

This gives

$$\begin{aligned} \det\left(x\left[\begin{array}{rr} 1 & 0 \\ 0 & 1 \end{array}\right] - \left[\begin{array}{rr} -5 & 2 \\ -7 & 4 \end{array}\right]\right) &= 0 \\ \det\left[\begin{array}{cc} x+5 & -2 \\ 7 & x-4 \end{array}\right] &= 0 \end{aligned}$$

Computing the determinant as usual, the result is

$$x^2 + x - 6 = 0$$

Solving this equation, we find that $\lambda_1 = 2$ and $\lambda_2 = -3$.

Now we need to find the basic eigenvectors for each λ. First we will find the eigenvectors for $\lambda_1 = 2$. We wish to find all vectors $X \neq 0$ such that $AX = 2X$. These are the solutions to $(2I - A)X = 0$.

$$\begin{aligned} \left(2\left[\begin{array}{rr} 1 & 0 \\ 0 & 1 \end{array}\right] - \left[\begin{array}{rr} -5 & 2 \\ -7 & 4 \end{array}\right]\right)\left[\begin{array}{c} x \\ y \end{array}\right] &= \left[\begin{array}{c} 0 \\ 0 \end{array}\right] \\ \left[\begin{array}{rr} 7 & -2 \\ 7 & -2 \end{array}\right]\left[\begin{array}{c} x \\ y \end{array}\right] &= \left[\begin{array}{c} 0 \\ 0 \end{array}\right] \end{aligned}$$

The augmented matrix for this system and corresponding reduced row-echelon form are given by

$$\left[\begin{array}{rr|r} 7 & -2 & 0 \\ 7 & -2 & 0 \end{array}\right] \rightarrow \cdots \rightarrow \left[\begin{array}{rr|r} 1 & -\frac{2}{7} & 0 \\ 0 & 0 & 0 \end{array}\right]$$

The solution is any vector of the form

$$\left[\begin{array}{c} \frac{2}{7}s \\ s \end{array}\right] = s\left[\begin{array}{c} \frac{2}{7} \\ 1 \end{array}\right]$$

Multiplying this vector by 7 we obtain a simpler description for the solution to this system, given by

$$t\left[\begin{array}{c} 2 \\ 7 \end{array}\right]$$

This gives the basic eigenvector for $\lambda_1 = 2$ as

$$\begin{bmatrix} 2 \\ 7 \end{bmatrix}$$

To check, we verify that $AX = 2X$ for this basic eigenvector.

$$\begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 7 \end{bmatrix} = \begin{bmatrix} 4 \\ 14 \end{bmatrix} = 2 \begin{bmatrix} 2 \\ 7 \end{bmatrix}$$

This is what we wanted, so we know this basic eigenvector is correct.

Next we will repeat this process to find the basic eigenvector for $\lambda_2 = -3$. We wish to find all vectors $X \neq 0$ such that $AX = -3X$. These are the solutions to $((-3)I - A)X = 0$.

$$\begin{aligned} \left((-3) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix} \right) \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} 2 & -2 \\ 7 & -7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{aligned}$$

The augmented matrix for this system and corresponding reduced row-echelon form are given by

$$\left[\begin{array}{rr|r} 2 & -2 & 0 \\ 7 & -7 & 0 \end{array} \right] \rightarrow \cdots \rightarrow \left[\begin{array}{rr|r} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array} \right]$$

The solution is any vector of the form

$$\begin{bmatrix} s \\ s \end{bmatrix} = s \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

This gives the basic eigenvector for $\lambda_2 = -3$ as

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

To check, we verify that $AX = -3X$ for this basic eigenvector.

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

This is what we wanted, so we know this basic eigenvector is correct. ◻

The following is an example using Procedure 6.5 for a 3×3 matrix.

Example 6.7: Find the Eigenvalues and Eigenvectors

Find the eigenvalues and eigenvectors for the matrix

$$A = \begin{bmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{bmatrix}$$

Solution. We will use Procedure 6.5. First we need to find the eigenvalues of A. Recall that they are the solutions of the equation

$$\det(xI - A) = 0$$

In this case the equation is

$$\det\left(x\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{bmatrix}\right) = 0$$

which becomes

$$\det\begin{bmatrix} x-5 & 10 & 5 \\ -2 & x-14 & -2 \\ 4 & 8 & x-6 \end{bmatrix} = 0$$

Using Laplace Expansion, compute this determinant and simplify. The result is the following equation.

$$(x-5)\left(x^2 - 20x + 100\right) = 0$$

Solving this equation, we find that the eigenvalues are $\lambda_1 = 5, \lambda_2 = 10$ and $\lambda_3 = 10$. Notice that 10 is a root of multiplicity two due to

$$x^2 - 20x + 100 = (x-10)^2$$

Therefore, $\lambda_2 = 10$ is an eigenvalue of multiplicity two.

Now that we have found the eigenvalues for A, we can compute the eigenvectors.

First we will find the basic eigenvectors for $\lambda_1 = 5$. In other words, we want to find all non-zero vectors X so that $AX = 5X$. This requires that we solve the equation $(5I - A)X = 0$ for X as follows.

$$\left(5\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{bmatrix}\right)\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

That is you need to find the solution to

$$\begin{bmatrix} 0 & 10 & 5 \\ -2 & -9 & -2 \\ 4 & 8 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

By now this is a familiar problem. You set up the augmented matrix and row reduce to get the solution. Thus the matrix you must row reduce is

$$\left[\begin{array}{rrr|r} 0 & 10 & 5 & 0 \\ -2 & -9 & -2 & 0 \\ 4 & 8 & -1 & 0 \end{array}\right]$$

The reduced row-echelon form is

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

and so the solution is any vector of the form

$$\left[\begin{array}{c} \frac{5}{4}s \\ -\frac{1}{2}s \\ s \end{array}\right] = s\left[\begin{array}{r} \frac{5}{4} \\ -\frac{1}{2} \\ 1 \end{array}\right]$$

where $s \in \mathbb{R}$. If we multiply this vector by 4, we obtain a simpler description for the solution to this system, as given by

$$t\left[\begin{array}{r} 5 \\ -2 \\ 4 \end{array}\right] \tag{6.3}$$

where $t \in \mathbb{R}$. Here, the basic eigenvector is given by

$$X_1 = \left[\begin{array}{r} 5 \\ -2 \\ 4 \end{array}\right]$$

Notice that we cannot let $t = 0$ here, because this would result in the zero vector and eigenvectors are never equal to 0! Other than this value, every other choice of t in 6.3 results in an eigenvector.

It is a good idea to check your work! To do so, we will take the original matrix and multiply by the basic eigenvector X_1. We check to see if we get $5X_1$.

$$\left[\begin{array}{rrr} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{array}\right]\left[\begin{array}{r} 5 \\ -2 \\ 4 \end{array}\right] = \left[\begin{array}{r} 25 \\ -10 \\ 20 \end{array}\right] = 5\left[\begin{array}{r} 5 \\ -2 \\ 4 \end{array}\right]$$

This is what we wanted, so we know that our calculations were correct.

Next we will find the basic eigenvectors for $\lambda_2, \lambda_3 = 10$. These vectors are the basic solutions to the equation,

$$\left(10\left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] - \left[\begin{array}{rrr} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{array}\right]\right)\left[\begin{array}{c} x \\ y \\ z \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \end{array}\right]$$

That is you must find the solutions to

$$\left[\begin{array}{rrr} 5 & 10 & 5 \\ -2 & -4 & -2 \\ 4 & 8 & 4 \end{array}\right]\left[\begin{array}{c} x \\ y \\ z \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \end{array}\right]$$

Consider the augmented matrix

$$\left[\begin{array}{rrr|r} 5 & 10 & 5 & 0 \\ -2 & -4 & -2 & 0 \\ 4 & 8 & 4 & 0 \end{array}\right]$$

The reduced row-echelon form for this matrix is

$$\left[\begin{array}{rrr|r} 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

and so the eigenvectors are of the form

$$\left[\begin{array}{c} -2s-t \\ s \\ t \end{array}\right] = s\left[\begin{array}{r} -2 \\ 1 \\ 0 \end{array}\right] + t\left[\begin{array}{r} -1 \\ 0 \\ 1 \end{array}\right]$$

Note that you can't pick t and s both equal to zero because this would result in the zero vector and eigenvectors are never equal to zero.

Here, there are two basic eigenvectors, given by

$$X_2 = \left[\begin{array}{r} -2 \\ 1 \\ 0 \end{array}\right], X_3 = \left[\begin{array}{r} -1 \\ 0 \\ 1 \end{array}\right]$$

Taking any (nonzero) linear combination of X_2 and X_3 will also result in an eigenvector for the eigenvalue $\lambda = 10$. As in the case for $\lambda = 5$, always check your work! For the first basic eigenvector, we can check $AX_2 = 10X_2$ as follows.

$$\left[\begin{array}{rrr} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{array}\right]\left[\begin{array}{r} -1 \\ 0 \\ 1 \end{array}\right] = \left[\begin{array}{r} -10 \\ 0 \\ 10 \end{array}\right] = 10\left[\begin{array}{r} -1 \\ 0 \\ 1 \end{array}\right]$$

This is what we wanted. Checking the second basic eigenvector, X_3, is left as an exercise. ◻

It is important to remember that for any eigenvector X, $X \neq 0$. However, it is possible to have eigenvalues equal to zero. This is illustrated in the following example.

Example 6.8: A Zero Eigenvalue

Let

$$A = \left[\begin{array}{rrr} 2 & 2 & -2 \\ 1 & 3 & -1 \\ -1 & 1 & 1 \end{array}\right]$$

Find the eigenvalues and eigenvectors of A.

Solution. First we find the eigenvalues of A. We will do so using Definition 6.2.

In order to find the eigenvalues of A, we solve the following equation.

$$\det(xI - A) = \det \begin{bmatrix} x-2 & -2 & 2 \\ -1 & x-3 & 1 \\ 1 & -1 & x-1 \end{bmatrix} = 0$$

This reduces to $x^3 - 6x^2 + 8x = 0$. You can verify that the solutions are $\lambda_1 = 0, \lambda_2 = 2, \lambda_3 = 4$. Notice that while eigenvectors can never equal 0, it is possible to have an eigenvalue equal to 0.

Now we will find the basic eigenvectors. For $\lambda_1 = 0$, we need to solve the equation $(0I - A)X = 0$. This equation becomes $-AX = 0$, and so the augmented matrix for finding the solutions is given by

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

The reduced row-echelon form is

$$\left[\begin{array}{rrr|r} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Therefore, the eigenvectors are of the form $t \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ where $t \neq 0$ and the basic eigenvector is given by

$$X_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

We can verify that this eigenvector is correct by checking that the equation $AX_1 = 0X_1$ holds. The product AX_1 is given by

$$AX_1 = \begin{bmatrix} 2 & 2 & -2 \\ 1 & 3 & -1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This clearly equals $0X_1$, so the equation holds. Hence, $AX_1 = 0X_1$ and so 0 is an eigenvalue of A.

Computing the other basic eigenvectors is left as an exercise. ◻

In the following sections, we examine ways to simplify this process of finding eigenvalues and eigenvectors by using properties of special types of matrices.

6.1.3. Eigenvalues and Eigenvectors for Special Types of Matrices

There are three special kinds of matrices which we can use to simplify the process of finding eigenvalues and eigenvectors. Throughout this section, we will discuss similar matrices, elementary matrices, as well as triangular matrices.

We begin with a definition.

Definition 6.9: Similar Matrices

Let A and B be $n \times n$ matrices. Suppose there exists an invertible matrix P such that

$$A = P^{-1}BP$$

*Then A and B are called **similar matrices**.*

It turns out that we can use the concept of similar matrices to help us find the eigenvalues of matrices. Consider the following lemma.

Lemma 6.10: Similar Matrices and Eigenvalues

Let A and B be similar matrices, so that $A = P^{-1}BP$ where A, B are $n \times n$ matrices and P is invertible. Then A, B have the same eigenvalues.

Proof. We need to show two things. First, we need to show that if $A = P^{-1}BP$, then A and B have the same eigenvalues. Secondly, we show that if A and B have the same eigenvalues, then $A = P^{-1}BP$.

Here is the proof of the first statement. Suppose $A = P^{-1}BP$ and λ is an eigenvalue of A, that is $AX = \lambda X$ for some $X \neq 0$. Then

$$P^{-1}BPX = \lambda X$$

and so

$$BPX = \lambda PX$$

Since P is one to one and $X \neq 0$, it follows that $PX \neq 0$. Here, PX plays the role of the eigenvector in this equation. Thus λ is also an eigenvalue of B. One can similarly verify that any eigenvalue of B is also an eigenvalue of A, and thus both matrices have the same eigenvalues as desired.

Proving the second statement is similar and is left as an exercise. □

Note that this proof also demonstrates that the eigenvectors of A and B will (generally) be *different*. We see in the proof that $AX = \lambda X$, while $B(PX) = \lambda (PX)$. Therefore, for an eigenvalue λ, A will have the eigenvector X while B will have the eigenvector PX.

The second special type of matrices we discuss in this section is elementary matrices. Recall from Definition 2.43 that an elementary matrix E is obtained by applying one row operation to the identity matrix.

It is possible to use elementary matrices to simplify a matrix before searching for its eigenvalues and eigenvectors. This is illustrated in the following example.

Example 6.11: Simplify Using Elementary Matrices

Find the eigenvalues for the matrix

$$A = \left[\begin{array}{rrr} 33 & 105 & 105 \\ 10 & 28 & 30 \\ -20 & -60 & -62 \end{array} \right]$$

Solution. This matrix has big numbers and therefore we would like to simplify as much as possible before computing the eigenvalues.

We will do so using row operations. First, add 2 times the second row to the third row. To do so, left multiply A by $E(2,2)$. Then right multiply A by the inverse of $E(2,2)$ as illustrated.

$$\left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{array} \right] \left[\begin{array}{rrr} 33 & 105 & 105 \\ 10 & 28 & 30 \\ -20 & -60 & -62 \end{array} \right] \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{array} \right] = \left[\begin{array}{rrr} 33 & -105 & 105 \\ 10 & -32 & 30 \\ 0 & 0 & -2 \end{array} \right]$$

By Lemma 6.10, the resulting matrix has the same eigenvalues as A where here, the matrix $E(2,2)$ plays the role of P.

We do this step again, as follows. In this step, we use the elementary matrix obtained by adding -3 times the second row to the first row.

$$\left[\begin{array}{rrr} 1 & -3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right] \left[\begin{array}{rrr} 33 & -105 & 105 \\ 10 & -32 & 30 \\ 0 & 0 & -2 \end{array} \right] \left[\begin{array}{rrr} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right] = \left[\begin{array}{rrr} 3 & 0 & 15 \\ 10 & -2 & 30 \\ 0 & 0 & -2 \end{array} \right] \tag{6.4}$$

Again by Lemma 6.10, this resulting matrix has the same eigenvalues as A. At this point, we can easily find the eigenvalues. Let

$$B = \left[\begin{array}{rrr} 3 & 0 & 15 \\ 10 & -2 & 30 \\ 0 & 0 & -2 \end{array} \right]$$

Then, we find the eigenvalues of B (and therefore of A) by solving the equation $\det(xI - B) = 0$. You should verify that this equation becomes

$$(x+2)(x+2)(x-3) = 0$$

Solving this equation results in eigenvalues of $\lambda_1 = -2, \lambda_2 = -2$, and $\lambda_3 = 3$. Therefore, these are also the eigenvalues of A.

$\square$

Through using elementary matrices, we were able to create a matrix for which finding the eigenvalues was easier than for A. At this point, you could go back to the original matrix A and solve $(\lambda I - A)X = 0$ to obtain the eigenvectors of A.

Notice that when you multiply on the right by an elementary matrix, you are doing the column operation defined by the elementary matrix. In 6.4 multiplication by the elementary matrix on the right merely involves taking three times the first column and adding to the second. Thus, without referring to the elementary matrices, the transition to the new matrix in 6.4 can be illustrated by

$$\left[\begin{array}{rrr} 33 & -105 & 105 \\ 10 & -32 & 30 \\ 0 & 0 & -2 \end{array}\right] \rightarrow \left[\begin{array}{rrr} 3 & -9 & 15 \\ 10 & -32 & 30 \\ 0 & 0 & -2 \end{array}\right] \rightarrow \left[\begin{array}{rrr} 3 & 0 & 15 \\ 10 & -2 & 30 \\ 0 & 0 & -2 \end{array}\right]$$

The third special type of matrix we will consider in this section is the triangular matrix. Recall Definition 3.12 which states that an upper (lower) triangular matrix contains all zeros below (above) the main diagonal. Remember that finding the determinant of a triangular matrix is a simple procedure of taking the product of the entries on the main diagonal.. It turns out that there is also a simple way to find the eigenvalues of a triangular matrix.

In the next example we will demonstrate that the eigenvalues of a triangular matrix are the entries on the main diagonal.

Example 6.12: Eigenvalues for a Triangular Matrix

Let $A = \left[\begin{array}{ccc} 1 & 2 & 4 \\ 0 & 4 & 7 \\ 0 & 0 & 6 \end{array}\right]$. *Find the eigenvalues of* A.

Solution. We need to solve the equation $\det(xI - A) = 0$ as follows

$$\det(xI - A) = \det \left[\begin{array}{ccc} x-1 & -2 & -4 \\ 0 & x-4 & -7 \\ 0 & 0 & x-6 \end{array}\right] = (x-1)(x-4)(x-6) = 0$$

Solving the equation $(x-1)(x-4)(x-6) = 0$ for x results in the eigenvalues $\lambda_1 = 1, \lambda_2 = 4$ and $\lambda_3 = 6$. Thus the eigenvalues are the entries on the main diagonal of the original matrix. ♢

The same result is true for lower triangular matrices. For any triangular matrix, the eigenvalues are equal to the entries on the main diagonal. To find the eigenvectors of a triangular matrix, we use the usual procedure.

In the next section, we explore an important process involving the eigenvalues and eigenvectors of a matrix.

Exercises

Exercise 6.1.1 *If A is an invertible $n \times n$ matrix, compare the eigenvalues of A and A^{-1}. More generally, for m an arbitrary integer, compare the eigenvalues of A and A^m.*

Exercise 6.1.2 *If A is an $n\times n$ matrix and c is a nonzero constant, compare the eigenvalues of A and cA.*

Exercise 6.1.3 *Let A, B be invertible $n \times n$ matrices which commute. That is, $AB = BA$. Suppose X is an eigenvector of B. Show that then AX must also be an eigenvector for B.*

Exercise 6.1.4 *Suppose A is an $n\times n$ matrix and it satisfies $A^m = A$ for some m a positive integer larger than 1. Show that if λ is an eigenvalue of A then $|\lambda|$ equals either 0 or 1.*

Exercise 6.1.5 *Show that if $AX = \lambda X$ and $AY = \lambda Y$, then whenever k, p are scalars,*

$$A(kX + pY) = \lambda (kX + pY)$$

Does this imply that $kX + pY$ is an eigenvector? Explain.

Exercise 6.1.6 *Suppose A is a 3×3 matrix and the following information is available.*

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

Find $A\begin{bmatrix} 1 \\ -4 \\ 3 \end{bmatrix}$.

Exercise 6.1.7 *Suppose A is a 3×3 matrix and the following information is available.*

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

Find $A\begin{bmatrix} 3 \\ -4 \\ 3 \end{bmatrix}$.

Exercise 6.1.8 *Suppose A is a 3×3 matrix and the following information is available.*

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

Find $A\begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix}$.

Exercise 6.1.9 *Find the eigenvalues and eigenvectors of the matrix*

$$\begin{bmatrix} -6 & -92 & 12 \\ 0 & 0 & 0 \\ -2 & -31 & 4 \end{bmatrix}$$

One eigenvalue is -2.

Exercise 6.1.10 *Find the eigenvalues and eigenvectors of the matrix*

$$\begin{bmatrix} -2 & -17 & -6 \\ 0 & 0 & 0 \\ 1 & 9 & 3 \end{bmatrix}$$

One eigenvalue is 1.

Exercise 6.1.11 *Find the eigenvalues and eigenvectors of the matrix*

$$\begin{bmatrix} 9 & 2 & 8 \\ 2 & -6 & -2 \\ -8 & 2 & -5 \end{bmatrix}$$

One eigenvalue is -3.

Exercise 6.1.12 *Find the eigenvalues and eigenvectors of the matrix*

$$\begin{bmatrix} 6 & 76 & 16 \\ -2 & -21 & -4 \\ 2 & 64 & 17 \end{bmatrix}$$

One eigenvalue is -2.

Exercise 6.1.13 *Find the eigenvalues and eigenvectors of the matrix*

$$\left[\begin{array}{rrr} 3 & 5 & 2 \\ -8 & -11 & -4 \\ 10 & 11 & 3 \end{array}\right]$$

One eigenvalue is -3.

Exercise 6.1.14 *Is it possible for a nonzero matrix to have only* 0 *as an eigenvalue?*

6.2 Diagonalization

Outcomes

A. *Determine when it is possible to diagonalize a matrix.*

B. *When possible, diagonalize a matrix.*

6.2.1. Diagonalizing a Matrix

The most important theorem about diagonalizability is the following major result.

Theorem 6.13: Eigenvectors and Diagonalizable Matrices

An $n \times n$ matrix A is diagonalizable if and only if there is an invertible matrix P given by

$$P = \left[\begin{array}{cccc} X_1 & X_2 & \cdots & X_n \end{array}\right]$$

where the X_k are eigenvectors of A.
Moreover if A is diagonalizable, the corresponding eigenvalues of A are the diagonal entries of the diagonal matrix D.

Proof. Suppose P is given as above as an invertible matrix whose columns are eigenvectors of A. Then P^{-1} is of the form

$$P^{-1} = \left[\begin{array}{c} W_1^T \\ W_2^T \\ \vdots \\ W_n^T \end{array}\right]$$

where $W_k^T X_j = \delta_{kj}$, which is the Kronecker's symbol defined by

$$\delta_{ij} = \left\{ \begin{array}{l} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{array} \right.$$

Then

$$
\begin{aligned}
P^{-1}AP &= \begin{bmatrix} W_1^T \\ W_2^T \\ \vdots \\ W_n^T \end{bmatrix} \begin{bmatrix} AX_1 & AX_2 & \cdots & AX_n \end{bmatrix} \\
&= \begin{bmatrix} W_1^T \\ W_2^T \\ \vdots \\ W_n^T \end{bmatrix} \begin{bmatrix} \lambda_1 X_1 & \lambda_2 X_2 & \cdots & \lambda_n X_n \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}
\end{aligned}
$$

Conversely, suppose A is diagonalizable so that $P^{-1}AP = D$. Let

$$P = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}$$

where the columns are the X_k and

$$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

Then

$$AP = PD = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

and so

$$\begin{bmatrix} AX_1 & AX_2 & \cdots & AX_n \end{bmatrix} = \begin{bmatrix} \lambda_1 X_1 & \lambda_2 X_2 & \cdots & \lambda_n X_n \end{bmatrix}$$

showing the X_k are eigenvectors of A and the λ_k are eigenvectors.
□

Notice that because the matrix P defined above is invertible it follows that the set of eigenvectors of A, $\{X_1, X_2, \cdots, X_n\}$, form a basis of $\mathbb{R}^n$.

We demonstrate the concept given in the above theorem in the next example. Note that not only are the columns of the matrix P formed by eigenvectors, but P must be invertible so must consist of a wide variety of eigenvectors. We achieve this by using basic eigenvectors for the columns of P.

Example 6.14: Diagonalize a Matrix

Let

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

Find an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$.

Solution. By Theorem 6.13 we use the eigenvectors of A as the columns of P, and the corresponding eigenvalues of A as the diagonal entries of D.

First, we will find the eigenvalues of A. To do so, we solve $\det(xI - A) = 0$ as follows.

$$\det\left(x \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 0 & 0 \\ 1 & 4 & -1 \\ -2 & -4 & 4 \end{bmatrix} \right) = 0$$

This computation is left as an exercise, and you should verify that the eigenvalues are $\lambda_1 = 2, \lambda_2 = 2$, and $\lambda_3 = 6$.

Next, we need to find the eigenvectors. We first find the eigenvectors for $\lambda_1, \lambda_2 = 2$. Solving $(2I - A)X = 0$ to find the eigenvectors, we find that the eigenvectors are

$$t \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

where t, s are scalars. Hence there are two basic eigenvectors which are given by

$$X_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, X_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

You can verify that the basic eigenvector for $\lambda_3 = 6$ is $X_3 = \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}$

Then, we construct the matrix P as follows.

$$P = \begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix} = \begin{bmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix}$$

That is, the columns of P are the basic eigenvectors of A. Then, you can verify that

$$P^{-1} = \begin{bmatrix} -\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{bmatrix}$$

Thus,

$$\begin{aligned}
P^{-1}AP &= \left[\begin{array}{rrr} -\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{array}\right] \left[\begin{array}{rrr} 2 & 0 & 0 \\ 1 & 4 & -1 \\ -2 & -4 & 4 \end{array}\right] \left[\begin{array}{rrr} -2 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{array}\right] \\
&= \left[\begin{array}{rrr} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{array}\right]
\end{aligned}$$

You can see that the result here is a diagonal matrix where the entries on the main diagonal are the eigenvalues of A. We expected this based on Theorem 6.13. Notice that eigenvalues on the main diagonal *must* be in the same order as the corresponding eigenvectors in P. ◻

Consider the next important theorem.

Theorem 6.15: Linearly Independent Eigenvectors

Let A be an $n \times n$ matrix, and suppose that A has distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$. For each i, let X_i be a λ_i-eigenvector of A. Then $\{X_1, X_2, \ldots, X_m\}$ is linearly independent.

The corollary that follows from this theorem gives a useful tool in determining if A is diagonalizable.

Corollary 6.16: Distinct Eigenvalues

Let A be an $n \times n$ matrix and suppose it has n distinct eigenvalues. Then it follows that A is diagonalizable.

It is possible that a matrix A cannot be diagonalized. In other words, we cannot find an invertible matrix P so that $P^{-1}AP = D$.

Consider the following example.

Example 6.17: A Matrix which cannot be Diagonalized

Let

$$A = \left[\begin{array}{rr} 1 & 1 \\ 0 & 1 \end{array}\right]$$

If possible, find an invertible matrix P and diagonal matrix D so that $P^{-1}AP = D$.

Solution. Through the usual procedure, we find that the eigenvalues of A are $\lambda_1 = 1, \lambda_2 = 1$. To find the eigenvectors, we solve the equation $(\lambda I - A) X = 0$. The matrix $(\lambda I - A)$ is given by

$$\left[\begin{array}{cc} \lambda - 1 & -1 \\ 0 & \lambda - 1 \end{array}\right]$$

Substituting in $\lambda = 1$, we have the matrix

$$\left[\begin{array}{cc} 1 - 1 & -1 \\ 0 & 1 - 1 \end{array}\right] = \left[\begin{array}{cc} 0 & -1 \\ 0 & 0 \end{array}\right]$$

Then, solving the equation $(\lambda I - A) X = 0$ involves carrying the following augmented matrix to its reduced row-echelon form.

$$\left[\begin{array}{cc|c} 0 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right] \rightarrow \cdots \rightarrow \left[\begin{array}{cc|c} 0 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

Then the eigenvectors are of the form

$$t \left[\begin{array}{c} 1 \\ 0 \end{array}\right]$$

and the basic eigenvector is

$$X_1 = \left[\begin{array}{c} 1 \\ 0 \end{array}\right]$$

In this case, the matrix A has one eigenvalue of multiplicity two, but only one basic eigenvector. In order to diagonalize A, we need to construct an invertible 2×2 matrix P. However, because A only has one basic eigenvector, we cannot construct this P. Notice that if we were to use X_1 as both columns of P, P would not be invertible. For this reason, we cannot repeat eigenvectors in P.

Hence this matrix cannot be diagonalized. ☐

The idea that a matrix may not be diagonalizable suggests that conditions exist to determine when it is possible to diagonalize a matrix. We saw earlier in Corollary 6.16 that an $n \times n$ matrix with n distinct eigenvalues is diagonalizable. It turns out that there are other useful diagonalizability tests.

First we need the following definition.

Definition 6.18: Eigenspace

Let A be an $n \times n$ matrix and $\lambda \in \mathbb{R}$. The eigenspace of A corresponding to λ, written $E_\lambda(A)$ is the set of all eigenvectors corresponding to λ.

In other words, the eigenspace $E_\lambda(A)$ is all X such that $AX = \lambda X$. Notice that this set can be written $E_\lambda(A) = \text{null}(\lambda I - A)$, showing that $E_\lambda(A)$ is a subspace of $\mathbb{R}^n$.

Recall that the multiplicity of an eigenvalue λ is the number of times that it occurs as a root of the characteristic polynomial.

Consider now the following lemma.

Lemma 6.19: Dimension of the Eigenspace

If A is an $n \times n$ matrix, then

$$\dim(E_\lambda(A)) \leq m$$

where λ is an eigenvalue of A of multiplicity m,

This result tells us that if λ is an eigenvalue of A, then the number of linearly independent λ-eigenvectors is never more than the multiplicity of λ. We now use this fact to provide a useful diagonalizability condition.

Theorem 6.20: Diagonalizability Condition

Let A be an $n \times n$ matrix A. Then A is diagonalizable if and only if for each eigenvalue λ of A, $\dim(E_\lambda(A))$ is equal to the multiplicity of λ.

Exercises

Exercise 6.2.1 *Find the eigenvalues and eigenvectors of the matrix*

$$\left[\begin{array}{rrr} 5 & -18 & -32 \\ 0 & 5 & 4 \\ 2 & -5 & -11 \end{array}\right]$$

One eigenvalue is 1. *Diagonalize if possible.*

Exercise 6.2.2 *Find the eigenvalues and eigenvectors of the matrix*

$$\left[\begin{array}{rrr} -13 & -28 & 28 \\ 4 & 9 & -8 \\ -4 & -8 & 9 \end{array}\right]$$

One eigenvalue is 3. *Diagonalize if possible.*

Exercise 6.2.3 *Find the eigenvalues and eigenvectors of the matrix*

$$\left[\begin{array}{rrr} 89 & 38 & 268 \\ 14 & 2 & 40 \\ -30 & -12 & -90 \end{array}\right]$$

One eigenvalue is -3. *Diagonalize if possible.*

Exercise 6.2.4 *Find the eigenvalues and eigenvectors of the matrix*

$$\left[\begin{array}{rrr} 1 & 90 & 0 \\ 0 & -2 & 0 \\ 3 & 89 & -2 \end{array}\right]$$

One eigenvalue is 1. *Diagonalize if possible.*

Exercise 6.2.5 *Find the eigenvalues and eigenvectors of the matrix*

$$\left[\begin{array}{rrr} 11 & 45 & 30 \\ 10 & 26 & 20 \\ -20 & -60 & -44 \end{array}\right]$$

One eigenvalue is 1*. Diagonalize if possible.*

Exercise 6.2.6 *Find the eigenvalues and eigenvectors of the matrix*

$$\left[\begin{array}{rrr} 95 & 25 & 24 \\ -196 & -53 & -48 \\ -164 & -42 & -43 \end{array}\right]$$

One eigenvalue is 5*. Diagonalize if possible.*

Exercise 6.2.7 *Suppose A is an $n \times n$ matrix and let V be an eigenvector such that $AV = \lambda V$. Also suppose the characteristic polynomial of A is*

$$\det(xI - A) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

Explain why

$$\left(A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I\right) V = 0$$

If A is diagonalizable, give a proof of the Cayley Hamilton theorem based on this. This theorem says A satisfies its characteristic equation,

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I = 0$$

Exercise 6.2.8 *Suppose the characteristic polynomial of an $n \times n$ matrix A is $1 - X^n$. Find A^{mn} where m is an integer.*

6.3 Applications of Spectral Theory

Outcomes

A. Use diagonalization to find a high power of a matrix.

B. Use diagonalization to solve dynamical systems.

6.3.1. Raising a Matrix to a High Power

Suppose we have a matrix A and we want to find A^{50}. One could try to multiply A with itself 50 times, but this is computationally extremely intensive (try it!). However diagonalization allows us to compute high powers of a matrix relatively easily. Suppose A is diagonalizable, so that $P^{-1}AP = D$. We can rearrange this equation to write $A = PDP^{-1}$.

Now, consider A^2. Since $A = PDP^{-1}$, it follows that

$$A^2 = \left(PDP^{-1}\right)^2 = PDP^{-1}PDP^{-1} = PD^2P^{-1}$$

Similarly,

$$A^3 = \left(PDP^{-1}\right)^3 = PDP^{-1}PDP^{-1}PDP^{-1} = PD^3P^{-1}$$

In general,

$$A^n = \left(PDP^{-1}\right)^n = PD^nP^{-1}$$

Therefore, we have reduced the problem to finding D^n. In order to compute D^n, then because D is diagonal we only need to raise every entry on the main diagonal of D to the power of n.

Through this method, we can compute large powers of matrices. Consider the following example.

Example 6.21: Raising a Matrix to a High Power

Let $A = \left[\begin{array}{rrr} 2 & 1 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{array}\right]$. *Find* A^{50}.

Solution. We will first diagonalize A. The steps are left as an exercise and you may wish to verify that the eigenvalues of A are $\lambda_1 = 1, \lambda_2 = 1$, and $\lambda_3 = 2$.

The basic eigenvectors corresponding to $\lambda_1, \lambda_2 = 1$ are

$$X_1 = \left[\begin{array}{r} 0 \\ 0 \\ 1 \end{array}\right], X_2 = \left[\begin{array}{r} -1 \\ 1 \\ 0 \end{array}\right]$$

The basic eigenvector corresponding to $\lambda_3 = 2$ is

$$X_3 = \left[\begin{array}{r} -1 \\ 0 \\ 1 \end{array}\right]$$

Now we construct P by using the basic eigenvectors of A as the columns of P. Thus

$$P = \left[\begin{array}{ccc} X_1 & X_2 & X_3 \end{array}\right] = \left[\begin{array}{rrr} 0 & -1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{array}\right]$$

Then also

$$P^{-1} = \left[\begin{array}{rrr} 1 & 1 & 1 \\ 0 & 1 & 0 \\ -1 & -1 & 0 \end{array} \right]$$

which you may wish to verify.

Then,

$$\begin{aligned} P^{-1}AP &= \left[\begin{array}{rrr} 1 & 1 & 1 \\ 0 & 1 & 0 \\ -1 & -1 & 0 \end{array} \right] \left[\begin{array}{rrr} 2 & 1 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{array} \right] \left[\begin{array}{rrr} 0 & -1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{array} \right] \\ &= \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{array} \right] \\ &= D \end{aligned}$$

Now it follows by rearranging the equation that

$$A = PDP^{-1} = \left[\begin{array}{rrr} 0 & -1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{array} \right] \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{array} \right] \left[\begin{array}{rrr} 1 & 1 & 1 \\ 0 & 1 & 0 \\ -1 & -1 & 0 \end{array} \right]$$

Therefore,

$$\begin{aligned} A^{50} &= PD^{50}P^{-1} \\ &= \left[\begin{array}{rrr} 0 & -1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{array} \right] \left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{array} \right]^{50} \left[\begin{array}{rrr} 1 & 1 & 1 \\ 0 & 1 & 0 \\ -1 & -1 & 0 \end{array} \right] \end{aligned}$$

By our discussion above, D^{50} is found as follows.

$$\left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{array} \right]^{50} = \left[\begin{array}{rrr} 1^{50} & 0 & 0 \\ 0 & 1^{50} & 0 \\ 0 & 0 & 2^{50} \end{array} \right]$$

It follows that

$$\begin{aligned} A^{50} &= \left[\begin{array}{rrr} 0 & -1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{array} \right] \left[\begin{array}{rrr} 1^{50} & 0 & 0 \\ 0 & 1^{50} & 0 \\ 0 & 0 & 2^{50} \end{array} \right] \left[\begin{array}{rrr} 1 & 1 & 1 \\ 0 & 1 & 0 \\ -1 & -1 & 0 \end{array} \right] \\ &= \left[\begin{array}{ccc} 2^{50} & -1+2^{50} & 0 \\ 0 & 1 & 0 \\ 1-2^{50} & 1-2^{50} & 1 \end{array} \right] \end{aligned}$$

□

Through diagonalization, we can efficiently compute a high power of A. Without this, we would be forced to multiply this by hand!

The next section explores another interesting application of diagonalization.

6.3.2. Raising a Symmetric Matrix to a High Power

We already have seen how to use matrix diagonalization to compute powers of matrices. This requires computing eigenvalues of the matrix A, and finding an invertible matrix of eigenvectors P such that $P^{-1}AP$ is diagonal. In this section we will see that if the matrix A is symmetric (see Definition 2.29), then we can actually find such a matrix P that is an orthogonal matrix of eigenvectors. Thus P^{-1} is simply its transpose P^T, and P^TAP is diagonal. When this happens we say that A is **orthogonally diagonalizable**

In fact this happens if and only if A is a symmetric matrix as shown in the following important theorem.

Theorem 6.22: Principal Axis Theorem

The following conditions are equivalent for an $n \times n$ matrix A:

1. *A is symmetric.*
2. *A has an orthonormal set of eigenvectors.*
3. *A is orthogonally diagonalizable.*

Proof. The complete proof is beyond this course, but to give an idea assume that A has an orthonormal set of eigenvectors, and let P consist of these eigenvectors as columns. Then $P^{-1} = P^T$, and $P^TAP = D$ a diagonal matrix. But then $A = PDP^T$, and

$$A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A$$

so A is symmetric.

Now given a symmetric matrix A, one shows that eigenvectors corresponding to different eigenvalues are always orthogonal. So it suffices to apply the Gram-Schmidt process on the set of basic eigenvectors of each eigenvalue to obtain an orthonormal set of eigenvectors. □

We demonstrate this in the following example.

Example 6.23: Orthogonal Diagonalization of a Symmetric Matrix

Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} \end{bmatrix}$. *Find an orthogonal matrix P such that P^TAP is a diagonal matrix.*

Solution.

In this case, verify that the eigenvalues are 2 and 1. First we will find an eigenvector for the eigenvalue 2. This involves row reducing the following augmented matrix.

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

The reduced row-echelon form is

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

and so an eigenvector is

$$\left[\begin{array}{c} 0 \\ 1 \\ 1 \end{array}\right]$$

Finally to obtain an eigenvector of length one (unit eigenvector) we simply divide this vector by its length to yield:

$$\left[\begin{array}{c} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{array}\right]$$

Next consider the case of the eigenvalue 1. To obtain basic eigenvectors, the matrix which needs to be row reduced in this case is

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

The reduced row-echelon form is

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

Therefore, the eigenvectors are of the form

$$\left[\begin{array}{c} s \\ -t \\ t \end{array}\right]$$

Note that all these vectors are automatically orthogonal to eigenvectors corresponding to the first eigenvalue. This follows from the fact that A is symmetric, as mentioned earlier.

We obtain basic eigenvectors

$$\left[\begin{array}{c} 1 \\ 0 \\ 0 \end{array}\right] \text{ and } \left[\begin{array}{r} 0 \\ -1 \\ 1 \end{array}\right]$$

Since they are themselves orthogonal (by luck here) we do not need to use the Gram-Schmidt process and instead simply normalize these vectors to obtain

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

An orthogonal matrix P to orthogonally diagonalize A is then obtained by letting these basic vectors be the columns.

$$P = \begin{bmatrix} 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}$$

We verify this works. $P^T AP$ is of the form

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

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

which is the desired diagonal matrix. □

We can now apply this technique to efficiently compute high powers of a symmetric matrix.

Example 6.24: Powers of a Symmetric Matrix

Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} \end{bmatrix}$. *Compute* A^7.

Solution. We found in Example 6.23 that $P^T AP = D$ is diagonal, where

$$P = \begin{bmatrix} 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \text{ and } D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Thus $A = PDP^T$ and $A^7 = PDP^T \; PDP^T \; \cdots PDP^T = PD^7P^T$ which gives:

$$A^7 = \begin{bmatrix} 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}^7 \begin{bmatrix} 0 & -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ 1 & 0 & 0 \\ 0 & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2^7 \end{bmatrix} \begin{bmatrix} 0 & -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ 1 & 0 & 0 \\ 0 & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 0 & -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ 1 & 0 & 0 \\ 0 & \frac{2^7}{2}\sqrt{2} & \frac{2^7}{2}\sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{2^7+1}{2} & \frac{2^7-1}{2} \\ 0 & \frac{2^7-1}{2} & \frac{2^7+1}{2} \end{bmatrix}$$

$\square$

6.3.3. Markov Matrices

There are applications which are of great importance which feature a special type of matrix. Matrices in which the columns are non-negative numbers which sum to one are called **Markov matrices**. An important application of Markov matrices is in population migration, as illustrated in the following definition.

Definition 6.25: Migration Matrices

*Let n locations be denoted by the numbers $1, 2, \cdots, n$. Suppose it is the case that each year the proportion of residents in location j which move to location i is a_{ij}. Also suppose no one escapes or emigrates from without these n locations. This last assumption requires $\sum_i a_{ij} = 1$, and means that the matrix A, such that $A = [a_{ij}]$, is a Markov matrix. In this context, A is also called a **migration matrix**.*

Consider the following example which demonstrates this situation.

Example 6.26: Migration Matrix

Let A be a Markov matrix given by

$$A = \left[\begin{array}{cc} .4 & .2 \\ .6 & .8 \end{array}\right]$$

Verify that A is a Markov matrix and describe the entries of A in terms of population migration.

Solution. The columns of A are comprised of non-negative numbers which sum to 1. Hence, A is a Markov matrix.

Now, consider the entries a_{ij} of A in terms of population. The entry $a_{11} = .4$ is the proportion of residents in location one which stay in location one in a given time period. Entry $a_{21} = .6$ is the proportion of residents in location 1 which move to location 2 in the same time period. Entry $a_{12} = .2$ is the proportion of residents in location 2 which move to location 1. Finally, entry $a_{22} = .8$ is the proportion of residents in location 2 which stay in location 2 in this time period.

Considered as a Markov matrix, these numbers are usually identified with probabilities. Hence, we can say that the probability that a resident of location one will stay in location one in the time period is .4. $\square$

Observe that in Example 6.26 if there was initially say 15 thousand people in location 1 and 10 thousands in location 2, then after one year there would be $.4 \times 15 + .2 \times 10 = 8$ thousands people in location 1 the following year, and similarly there would be $.6 \times 15 + .8 \times 10 = 17$ thousands people in location 2 the following year.

More generally let $X_n = [x_{1n} \cdots x_{mn}]^T$ where x_{in} is the population of location i at time period n. We call X_n the **state vector at period** n. In particular, we call X_0 the initial state vector. Letting A be the migration matrix, we compute the population in each location i one time period later by AX_n. In order to find the population of location i after k years, we compute the i^{th} component of $A^k X$. This discussion is summarized in the following theorem.

Theorem 6.27: State Vector

Let A be the migration matrix of a population and let X_n be the vector whose entries give the population of each location at time period n. Then X_n is the state vector at period n and it follows that

$$X_{n+1} = AX_n$$

The sum of the entries of X_n will equal the sum of the entries of the initial vector X_0. Since the columns of A sum to 1, this sum is preserved for every multiplication by A as demonstrated below.

$$\sum_i \sum_j a_{ij} x_j = \sum_j x_j \left(\sum_i a_{ij} \right) = \sum_j x_j$$

Consider the following example.

Example 6.28: Using a Migration Matrix

Consider the migration matrix

$$A = \left[\begin{array}{rrr} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{array} \right]$$

for locations $1, 2,$ *and* 3*. Suppose initially there are* 100 *residents in location* 1*,* 200 *in location* 2 *and* 400 *in location* 3*. Find the population in the three locations after* $1, 2,$ *and* 10 *units of time.*

Solution. Using Theorem 6.27 we can find the population in each location using the equation $X_{n+1} = AX_n$. For the population after 1 unit, we calculate $X_1 = AX_0$ as follows.

$$\begin{array}{rcl} X_1 & = & AX_0 \\ \left[\begin{array}{c} x_{11} \\ x_{21} \\ x_{31} \end{array} \right] & = & \left[\begin{array}{rrr} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{array} \right] \left[\begin{array}{c} 100 \\ 200 \\ 400 \end{array} \right] \\ & = & \left[\begin{array}{c} 100 \\ 180 \\ 420 \end{array} \right] \end{array}$$

Therefore after one time period, location 1 has 100 residents, location 2 has 180, and location 3 has 420. Notice that the **total** population is unchanged, it simply migrates within the given locations. We find the locations after two time periods in the same way.

$$\begin{array}{rcl} X_2 & = & AX_1 \\ \left[\begin{array}{c} x_{12} \\ x_{22} \\ x_{32} \end{array} \right] & = & \left[\begin{array}{rrr} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{array} \right] \left[\begin{array}{c} 100 \\ 180 \\ 420 \end{array} \right] \\ & = & \left[\begin{array}{c} 102 \\ 164 \\ 434 \end{array} \right] \end{array}$$

We could progress in this manner to find the populations after 10 time periods. However from our above discussion, we can simply calculate $(A^n X_0)_i$, where n denotes the number of time periods which have passed. Therefore, we compute the populations in each location after 10 units of time as follows.

$$\begin{array}{rcl} X_{10} & = & A^{10} X_0 \\ \left[\begin{array}{c} x_{110} \\ x_{210} \\ x_{310} \end{array} \right] & = & \left[\begin{array}{rrr} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{array} \right]^{10} \left[\begin{array}{c} 100 \\ 200 \\ 400 \end{array} \right] \\ & = & \left[\begin{array}{c} 115.085\,829\,22 \\ 120.130\,672\,44 \\ 464.783\,498\,34 \end{array} \right] \end{array}$$

Since we are speaking about populations, we would need to round these numbers to provide a logical answer. Therefore, we can say that after 10 units of time, there will be 115 residents in location one, 120 in location two, and 465 in location three.

□

Suppose we wish to know how many residents will be in a certain location after a very long time. It turns out that if some power of the migration matrix has all positive entries, then there is a vector X_s such that $A^n X_0$ approaches X_s as n becomes very large. Hence as more time passes and n increases, $A^n X_0$ will become closer to the vector X_s.

Consider Theorem 6.27. Let n increase so that X_n approaches X_s. As X_n becomes closer to X_s, so too does X_{n+1}. For sufficiently large n, the statement $X_{n+1} = AX_n$ can be written as $X_s = AX_s$.

This discussion motivates the following theorem.

Theorem 6.29: Steady State Vector

Let A be a migration matrix. Then there exists a ***steady state vector*** *written X_s such that*

$$X_s = AX_s$$

where X_s has positive entries which have the same sum as the entries of X_0.
As n increases, the state vectors X_n will approach X_s.

Note that the condition in Theorem 6.29 can be written as $(I - A)X_s = 0$, representing a homogeneous system of equations.

Consider the following example. Notice that it is the same example as the Example 6.28 but here it will involve a longer time frame.

Example 6.30: Populations over the Long Run

Consider the migration matrix

$$A = \begin{bmatrix} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{bmatrix}$$

for locations $1, 2,$ and 3. Suppose initially there are 100 residents in location 1, 200 in location 2 and 400 in location 4. Find the population in the three locations after a long time.

Solution. By Theorem 6.29 the steady state vector X_s can be found by solving the system $(I - A)X_s = 0$.

Thus we need to find a solution to

$$\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{bmatrix} \right) \begin{bmatrix} x_{1s} \\ x_{2s} \\ x_{3s} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The augmented matrix and the resulting reduced row-echelon form are given by

$$\left[\begin{array}{rrr|r} 0.4 & 0 & -0.1 & 0 \\ -0.2 & 0.2 & 0 & 0 \\ -0.2 & -0.2 & 0.1 & 0 \end{array}\right] \rightarrow \cdots \rightarrow \left[\begin{array}{rrr|r} 1 & 0 & -0.25 & 0 \\ 0 & 1 & -0.25 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Therefore, the eigenvectors are

$$t \left[\begin{array}{c} 0.25 \\ 0.25 \\ 1 \end{array}\right]$$

The initial vector X_0 is given by

$$\left[\begin{array}{c} 100 \\ 200 \\ 400 \end{array}\right]$$

Now all that remains is to choose the value of t such that

$$0.25t + 0.25t + t = 100 + 200 + 400$$

Solving this equation for t yields $t = \frac{1400}{3}$. Therefore the population in the long run is given by

$$\frac{1400}{3}\left[\begin{array}{c} 0.25 \\ 0.25 \\ 1 \end{array}\right] = \left[\begin{array}{c} 116.666\,666\,666\,666\,7 \\ 116.666\,666\,666\,666\,7 \\ 466.666\,666\,666\,666\,7 \end{array}\right]$$

Again, because we are working with populations, these values need to be rounded. The steady state vector X_s is given by

$$\left[\begin{array}{c} 117 \\ 117 \\ 466 \end{array}\right]$$

□

We can see that the numbers we calculated in Example 6.28 for the populations after the 10^{th} unit of time are not far from the long term values.

Consider another example.

Example 6.31: Populations After a Long Time

Suppose a migration matrix is given by

$$A = \left[\begin{array}{ccc} \frac{1}{5} & \frac{1}{2} & \frac{1}{5} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{11}{20} & \frac{1}{4} & \frac{3}{10} \end{array}\right]$$

Find the comparison between the populations in the three locations after a long time.

Solution. In order to compare the populations in the long term, we want to find the steady state vector X_s. Solve

$$\left(\left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right] - \left[\begin{array}{ccc} \frac{1}{5} & \frac{1}{2} & \frac{1}{5} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{11}{20} & \frac{1}{4} & \frac{3}{10} \end{array} \right] \right) \left[\begin{array}{c} x_{1s} \\ x_{2s} \\ x_{3s} \end{array} \right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right]$$

The augmented matrix and the resulting reduced row-echelon form are given by

$$\left[\begin{array}{rrr|r} \frac{4}{5} & -\frac{1}{2} & -\frac{1}{5} & 0 \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{2} & 0 \\ -\frac{11}{20} & -\frac{1}{4} & \frac{7}{10} & 0 \end{array} \right] \rightarrow \cdots \rightarrow \left[\begin{array}{rrr|r} 1 & 0 & -\frac{16}{19} & 0 \\ 0 & 1 & -\frac{18}{19} & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

and so an eigenvector is

$$\left[\begin{array}{c} 16 \\ 18 \\ 19 \end{array} \right]$$

Therefore, the proportion of population in location 2 to location 1 is given by $\frac{18}{16}$. The proportion of population 3 to location 2 is given by $\frac{19}{18}$.

$\square$

Eigenvalues of Markov Matrices

The following is an important proposition.

Proposition 6.32: Eigenvalues of a Migration Matrix

Let $A = [a_{ij}]$ be a migration matrix. Then 1 is always an eigenvalue for A.

Proof. Remember that the determinant of a matrix always equals that of its transpose. Therefore,

$$\det(xI - A) = \det\left((xI - A)^T\right) = \det\left(xI - A^T\right)$$

because $I^T = I$. Thus the characteristic equation for A is the same as the characteristic equation for A^T. Consequently, A and A^T have the same eigenvalues. We will show that 1 is an eigenvalue for A^T and then it will follow that 1 is an eigenvalue for A.

Remember that for a migration matrix, $\sum_i a_{ij} = 1$. Therefore, if $A^T = [b_{ij}]$ with $b_{ij} = a_{ji}$, it follows that

$$\sum_j b_{ij} = \sum_j a_{ji} = 1$$

Therefore, from matrix multiplication,

$$A^T \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \sum_j b_{ij} \\ \vdots \\ \sum_j b_{ij} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

Notice that this shows that $\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ is an eigenvector for A^T corresponding to the eigenvalue, $\lambda = 1$. As explained above, this shows that $\lambda = 1$ is an eigenvalue for A because A and A^T have the same eigenvalues. $\square$

6.3.4. Dynamical Systems

The migration matrices discussed above give an example of a discrete dynamical system. We call them discrete because they involve discrete values taken at a sequence of points rather than on a continuous interval of time.

An example of a situation which can be studied in this way is a predator prey model. Consider the following model where x is the number of prey and y the number of predators in a certain area at a certain time. These are functions of $n \in \mathbb{N}$ where $n = 1, 2, \cdots$ are the ends of intervals of time which may be of interest in the problem. In other words, $x(n)$ is the number of prey at the end of the n^{th} interval of time. An example of this situation may be modeled by the following equation

$$\begin{bmatrix} x(n+1) \\ y(n+1) \end{bmatrix} = \begin{bmatrix} 2 & -3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x(n) \\ y(n) \end{bmatrix}$$

This says that from time period n to $n+1$, x increases if there are more x and decreases as there are more y. In the context of this example, this means that as the number of predators increases, the number of prey decreases. As for y, it increases if there are more y and also if there are more x.

This is an example of a matrix recurrence which we define now.

Definition 6.33: Matrix Recurrence

Suppose a dynamical system is given by

$$\begin{aligned} x_{n+1} &= ax_n + by_n \\ y_{n+1} &= cx_n + dy_n \end{aligned}$$

This system can be expressed as $V_{n+1} = AV_n$ *where* $V_n = \begin{bmatrix} x_n \\ y_n \end{bmatrix}$ *and* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

In this section, we will examine how to find solutions to a dynamical system given certain initial conditions. This process involves several concepts previously studied, including matrix diagonalization and Markov matrices. The procedure is given as follows. Recall that when diagonalized, we can write $A^n = PD^nP^{-1}$.

Procedure 6.34: Solving a Dynamical System

Suppose a dynamical system is given by

$$\begin{array}{rcl} x_{n+1} & = & ax_n + by_n \\ y_{n+1} & = & cx_n + dy_n \end{array}$$

Given initial conditions x_0 and y_0, the solutions to the system are found as follows:

1. *Express the dynamical system in the form $V_{n+1} = AV_n$.*
2. *Diagonalize A to be written as $A = PDP^{-1}$.*
3. *Then $V_n = PD^nP^{-1}V_0$ where V_0 is the vector containing the initial conditions.*
4. *If given specific values for n, substitute into this equation. Otherwise, find a general solution for n.*

We will now consider an example in detail.

Example 6.35: Solutions of a Discrete Dynamical System

Suppose a dynamical system is given by

$$\begin{array}{rcl} x_{n+1} & = & 1.5x_n - 0.5y_n \\ y_{n+1} & = & 1.0x_n \end{array}$$

Express this system as a matrix recurrence and find solutions to the dynamical system for initial conditions $x_0 = 20, y_0 = 10$.

Solution. First, we express the system as a matrix recurrence.

$$\begin{array}{rcl} V_{n+1} & = & AV_n \\ \left[\begin{array}{c} x(n+1) \\ y(n+1) \end{array}\right] & = & \left[\begin{array}{rr} 1.5 & -0.5 \\ 1.0 & 0 \end{array}\right] \left[\begin{array}{c} x(n) \\ y(n) \end{array}\right] \end{array}$$

Then

$$A = \left[\begin{array}{rr} 1.5 & -0.5 \\ 1.0 & 0 \end{array}\right]$$

You can verify that the eigenvalues of A are 1 and .5. By diagonalizing, we can write A in the form

$$P^{-1}DP = \left[\begin{array}{rr} 1 & 1 \\ 1 & 2 \end{array}\right] \left[\begin{array}{rr} 1 & 0 \\ 0 & .5 \end{array}\right] \left[\begin{array}{rr} 2 & -1 \\ -1 & 1 \end{array}\right]$$

Now given an initial condition

$$V_0 = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$$

the solution to the dynamical system is given by

$$\begin{aligned} V_n &= PD^nP^{-1}V_0 \\ \begin{bmatrix} x(n) \\ y(n) \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & .5 \end{bmatrix}^n \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (.5)^n \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \\ &= \begin{bmatrix} y_0((.5)^n - 1) - x_0((.5)^n - 2) \\ y_0(2(.5)^n - 1) - x_0(2(.5)^n - 2) \end{bmatrix} \end{aligned}$$

If we let n become arbitrarily large, this vector approaches

$$\begin{bmatrix} 2x_0 - y_0 \\ 2x_0 - y_0 \end{bmatrix}$$

Thus for large n,

$$\begin{bmatrix} x(n) \\ y(n) \end{bmatrix} \approx \begin{bmatrix} 2x_0 - y_0 \\ 2x_0 - y_0 \end{bmatrix}$$

Now suppose the initial condition is given by

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 20 \\ 10 \end{bmatrix}$$

Then, we can find solutions for various values of n. Here are the solutions for values of n between 1 and 5

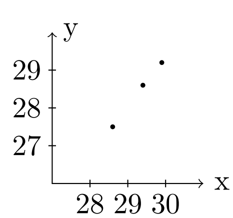
$$n = 1 : \begin{bmatrix} 25.0 \\ 20.0 \end{bmatrix}, n = 2 : \begin{bmatrix} 27.5 \\ 25.0 \end{bmatrix}, n = 3 : \begin{bmatrix} 28.75 \\ 27.5 \end{bmatrix}$$

$$n = 4 : \begin{bmatrix} 29.375 \\ 28.75 \end{bmatrix}, n = 5 : \begin{bmatrix} 29.688 \\ 29.375 \end{bmatrix}$$

Notice that as n increases, we approach the vector given by

$$\begin{bmatrix} 2x_0 - y_0 \\ 2x_0 - y_0 \end{bmatrix} = \begin{bmatrix} 2(20) - 10 \\ 2(20) - 10 \end{bmatrix} = \begin{bmatrix} 30 \\ 30 \end{bmatrix}$$

These solutions are graphed in the following figure.



$\square$

The following example demonstrates another system which exhibits some interesting behavior. When we graph the solutions, it is possible for the ordered pairs to spiral around the origin.

Example 6.36: Finding Solutions to a Dynamical System

Suppose a dynamical system is of the form

$$\left[\begin{array}{c} x(n+1) \\ y(n+1) \end{array}\right] = \left[\begin{array}{rr} 0.7 & 0.7 \\ -0.7 & 0.7 \end{array}\right] \left[\begin{array}{c} x(n) \\ y(n) \end{array}\right]$$

Find solutions to the dynamical system for given initial conditions.

Solution. Let

$$A = \left[\begin{array}{rr} 0.7 & 0.7 \\ -0.7 & 0.7 \end{array}\right]$$

To find solutions, we must diagonalize A. You can verify that the eigenvalues of A are complex and are given by $\lambda_1 = .7 + .7i$ and $\lambda_2 = .7 - .7i$. The eigenvector for $\lambda_1 = .7 + .7i$ is

$$\left[\begin{array}{c} 1 \\ i \end{array}\right]$$

and that the eigenvector for $\lambda_2 = .7 - .7i$ is

$$\left[\begin{array}{r} 1 \\ -i \end{array}\right]$$

Thus the matrix A can be written in the form

$$\left[\begin{array}{rr} 1 & 1 \\ i & -i \end{array}\right] \left[\begin{array}{cc} .7+.7i & 0 \\ 0 & .7-.7i \end{array}\right] \left[\begin{array}{rr} \frac{1}{2} & -\frac{1}{2}i \\ \frac{1}{2} & \frac{1}{2}i \end{array}\right]$$

and so,

$$\begin{aligned} V_n &= PD^nP^{-1}V_0 \\ \left[\begin{array}{c} x(n) \\ y(n) \end{array}\right] &= \left[\begin{array}{rr} 1 & 1 \\ i & -i \end{array}\right] \left[\begin{array}{cc} (.7+.7i)^n & 0 \\ 0 & (.7-.7i)^n \end{array}\right] \left[\begin{array}{rr} \frac{1}{2} & -\frac{1}{2}i \\ \frac{1}{2} & \frac{1}{2}i \end{array}\right] \left[\begin{array}{c} x_0 \\ y_0 \end{array}\right] \end{aligned}$$

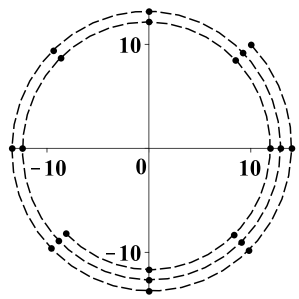
The explicit solution is given by

$$\left[\begin{array}{c} x_0\left(\frac{1}{2}(0.7-0.7i)^n + \frac{1}{2}(0.7+0.7i)^n\right) + y_0\left(\frac{1}{2}i(0.7-0.7i)^n - \frac{1}{2}i(0.7+0.7i)^n\right) \\ y_0\left(\frac{1}{2}(0.7-0.7i)^n + \frac{1}{2}(0.7+0.7i)^n\right) - x_0\left(\frac{1}{2}i(0.7-0.7i)^n - \frac{1}{2}i(0.7+0.7i)^n\right) \end{array}\right]$$

Suppose the initial condition is

$$\left[\begin{array}{c} x_0 \\ y_0 \end{array}\right] = \left[\begin{array}{c} 10 \\ 10 \end{array}\right]$$

Then one obtains the following sequence of values which are graphed below by letting $n = 1, 2, \cdots, 20$



In this picture, the dots are the values and the dashed line is to help to picture what is happening.

These points are getting gradually closer to the origin, but they are circling the origin in the clockwise direction as they do so. As n increases, the vector $\begin{bmatrix} x(n) \\ y(n) \end{bmatrix}$ approaches $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\square$

This type of behavior along with complex eigenvalues is typical of the deviations from an equilibrium point in the Lotka Volterra system of differential equations which is a famous model for predator-prey interactions. These differential equations are given by

$$\begin{aligned} x' &= x(a - by) \\ y' &= -y(c - dx) \end{aligned}$$

where a, b, c, d are positive constants. For example, you might have X be the population of moose and Y the population of wolves on an island.

Note that these equations make logical sense. The top says that the rate at which the moose population increases would be aX if there were no predators Y. However, this is modified by multiplying instead by $(a - bY)$ because if there are predators, these will militate against the population of moose. The more predators there are, the more pronounced is this effect. As to the predator equation, you can see that the equations predict that if there are many prey around, then the rate of growth of the predators would seem to be high. However, this is modified by the term $-cY$ because if there are many predators, there would be competition for the available food supply and this would tend to decrease Y'.

The behavior near an equilibrium point, which is a point where the right side of the differential equations equals zero, is of great interest. In this case, the equilibrium point is

$$x = \frac{c}{d}, y = \frac{a}{b}$$

Then one defines new variables according to the formula

$$x + \frac{c}{d} = x,\ y = y + \frac{a}{b}$$

In terms of these new variables, the differential equations become

$$\begin{aligned} x' &= \left(x+\frac{c}{d}\right)\left(a-b\left(y+\frac{a}{b}\right)\right) \\ y' &= -\left(y+\frac{a}{b}\right)\left(c-d\left(x+\frac{c}{d}\right)\right) \end{aligned}$$

Multiplying out the right sides yields

$$\begin{aligned} x' &= -bxy-b\frac{c}{d}y \\ y' &= dxy+\frac{a}{b}dx \end{aligned}$$

The interest is for x, y small and so these equations are essentially equal to

$$x' = -b\frac{c}{d}y, \; y' = \frac{a}{b}dx$$

Replace x' with the difference quotient $\frac{x(t+h)-x(t)}{h}$ where h is a small positive number and y' with a similar difference quotient. For example one could have h correspond to one day or even one hour. Thus, for h small enough, the following would seem to be a good approximation to the differential equations.

$$\begin{aligned} x(t+h) &= x(t)-hb\frac{c}{d}y \\ y(t+h) &= y(t)+h\frac{a}{b}dx \end{aligned}$$

Let $1, 2, 3, \cdots$ denote the ends of discrete intervals of time having length h chosen above. Then the above equations take the form

$$\left[\begin{array}{c} x(n+1) \\ y(n+1) \end{array}\right] = \left[\begin{array}{cc} 1 & -\frac{hbc}{d} \\ \frac{had}{b} & 1 \end{array}\right]\left[\begin{array}{c} x(n) \\ y(n) \end{array}\right]$$

Note that the eigenvalues of this matrix are always complex.

We are not interested in time intervals of length h for h very small. Instead, we are interested in much longer lengths of time. Thus, replacing the time interval with mh,

$$\left[\begin{array}{c} x(n+m) \\ y(n+m) \end{array}\right] = \left[\begin{array}{cc} 1 & -\frac{hbc}{d} \\ \frac{had}{b} & 1 \end{array}\right]^m\left[\begin{array}{c} x(n) \\ y(n) \end{array}\right]$$

For example, if $m = 2$, you would have

$$\left[\begin{array}{c} x(n+2) \\ y(n+2) \end{array}\right] = \left[\begin{array}{cc} 1-ach^2 & -2b\frac{c}{d}h \\ 2\frac{a}{b}dh & 1-ach^2 \end{array}\right]\left[\begin{array}{c} x(n) \\ y(n) \end{array}\right]$$

Note that most of the time, the eigenvalues of the new matrix will be complex.

You can also notice that the upper right corner will be negative by considering higher powers of the matrix. Thus letting $1, 2, 3, \cdots$ denote the ends of discrete intervals of time, the desired discrete dynamical system is of the form

$$\left[\begin{array}{c} x(n+1) \\ y(n+1) \end{array}\right] = \left[\begin{array}{cc} a & -b \\ c & d \end{array}\right] \left[\begin{array}{c} x(n) \\ y(n) \end{array}\right]$$

where a, b, c, d are positive constants and the matrix will likely have complex eigenvalues because it is a power of a matrix which has complex eigenvalues.

You can see from the above discussion that if the eigenvalues of the matrix used to define the dynamical system are less than 1 in absolute value, then the origin is stable in the sense that as $n \to \infty$, the solution converges to the origin. If either eigenvalue is larger than 1 in absolute value, then the solutions to the dynamical system will usually be unbounded, unless the initial condition is chosen very carefully. The next example exhibits the case where one eigenvalue is larger than 1 and the other is smaller than 1.

The following example demonstrates a familiar concept as a dynamical system.

Example 6.37: The Fibonacci Sequence

The Fibonacci sequence is the sequence given by

$$1, 1, 2, 3, 5, \cdots$$

which is defined recursively in the form

$$x(0) = 1 = x(1), \ x(n+2) = x(n+1) + x(n)$$

Show how the Fibonacci Sequence can be considered a dynamical system.

Solution. This sequence is extremely important in the study of reproducing rabbits. It can be considered as a dynamical system as follows. Let $y(n) = x(n+1)$. Then the above recurrence relation can be written as

$$\left[\begin{array}{c} x(n+1) \\ y(n+1) \end{array}\right] = \left[\begin{array}{cc} 0 & 1 \\ 1 & 1 \end{array}\right] \left[\begin{array}{c} x(n) \\ y(n) \end{array}\right], \ \left[\begin{array}{c} x(0) \\ y(0) \end{array}\right] = \left[\begin{array}{c} 1 \\ 1 \end{array}\right]$$

Let

$$A = \left[\begin{array}{cc} 0 & 1 \\ 1 & 1 \end{array}\right]$$

The eigenvalues of the matrix A are $\lambda_1 = \frac{1}{2} - \frac{1}{2}\sqrt{5}$ and $\lambda_2 = \frac{1}{2}\sqrt{5} + \frac{1}{2}$. The corresponding eigenvectors are, respectively,

$$X_1 = \left[\begin{array}{c} -\frac{1}{2}\sqrt{5} - \frac{1}{2} \\ 1 \end{array}\right], X_2 = \left[\begin{array}{c} \frac{1}{2}\sqrt{5} - \frac{1}{2} \\ 1 \end{array}\right]$$

You can see from a short computation that one of the eigenvalues is smaller than 1 in absolute value while the other is larger than 1 in absolute value. Now, diagonalizing A gives us

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

$$= \begin{bmatrix} \frac{1}{2}\sqrt{5}+\frac{1}{2} & 0 \\ 0 & \frac{1}{2}-\frac{1}{2}\sqrt{5} \end{bmatrix}$$

Then it follows that for a given initial condition, the solution to this dynamical system is of the form

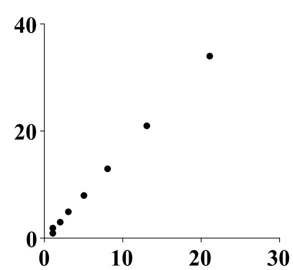
$$\begin{bmatrix} x(n) \\ y(n) \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\sqrt{5}-\frac{1}{2} & -\frac{1}{2}\sqrt{5}-\frac{1}{2} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \left(\frac{1}{2}\sqrt{5}+\frac{1}{2}\right)^n & 0 \\ 0 & \left(\frac{1}{2}-\frac{1}{2}\sqrt{5}\right)^n \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{5}\sqrt{5} & \frac{1}{10}\sqrt{5}+\frac{1}{2} \\ -\frac{1}{5}\sqrt{5} & \frac{1}{5}\sqrt{5}\left(\frac{1}{2}\sqrt{5}-\frac{1}{2}\right) \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

It follows that

$$x(n) = \left(\frac{1}{2}\sqrt{5}+\frac{1}{2}\right)^n \left(\frac{1}{10}\sqrt{5}+\frac{1}{2}\right) + \left(\frac{1}{2}-\frac{1}{2}\sqrt{5}\right)^n \left(\frac{1}{2}-\frac{1}{10}\sqrt{5}\right)$$

$\square$

Here is a picture of the ordered pairs $(x(n), y(n))$ for $n = 0, 1, \cdots, n$.



There is so much more that can be said about dynamical systems. It is a major topic of study in differential equations and what is given above is just an introduction.

Exercises

Exercise 6.3.1 *Let* $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. *Diagonalize* A *to find* A^{10}.

Exercise 6.3.2 *Let* $A = \begin{bmatrix} 1 & 4 & 1 \\ 0 & 2 & 5 \\ 0 & 0 & 5 \end{bmatrix}$. *Diagonalize* A *to find* A^{50}.

Exercise 6.3.3 *Let* $A = \begin{bmatrix} 1 & -2 & -1 \\ 2 & -1 & 1 \\ -2 & 3 & 1 \end{bmatrix}$. *Diagonalize* A *to find* A^{100}.

Exercise 6.3.4 *The following is a Markov (migration) matrix for three locations*

$$\begin{bmatrix} \frac{7}{10} & \frac{1}{9} & \frac{1}{5} \\ \frac{1}{10} & \frac{7}{9} & \frac{2}{5} \\ \frac{1}{5} & \frac{1}{9} & \frac{2}{5} \end{bmatrix}$$

1. *Initially, there are* 90 *people in location* 1*,* 81 *in location* 2*, and* 85 *in location* 3*. How many are in each location after one time period?*

2. *The total number of individuals in the migration process is* 256*. After a long time, how many are in each location?*

Exercise 6.3.5 *The following is a Markov (migration) matrix for three locations*

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

1. *Initially, there are* 130 *individuals in location* 1*,* 300 *in location* 2*, and* 70 *in location* 3*. How many are in each location after two time periods?*

2. *The total number of individuals in the migration process is* 500*. After a long time, how many are in each location?*

Exercise 6.3.6 *The following is a Markov (migration) matrix for three locations*

$$\begin{bmatrix} \frac{3}{10} & \frac{3}{8} & \frac{1}{3} \\ \frac{1}{10} & \frac{3}{8} & \frac{1}{3} \\ \frac{3}{5} & \frac{1}{4} & \frac{1}{3} \end{bmatrix}$$

The total number of individuals in the migration process is 480*. After a long time, how many are in each location?*

Exercise 6.3.7 *The following is a Markov (migration) matrix for three locations*

$$\begin{bmatrix} \frac{3}{10} & \frac{1}{3} & \frac{1}{5} \\ \frac{3}{10} & \frac{1}{3} & \frac{7}{10} \\ \frac{2}{5} & \frac{1}{3} & \frac{1}{10} \end{bmatrix}$$

The total number of individuals in the migration process is 1155*. After a long time, how many are in each location?*

Exercise 6.3.8 *The following is a Markov (migration) matrix for three locations*

$$\begin{bmatrix} \frac{2}{5} & \frac{1}{10} & \frac{1}{8} \\ \frac{3}{10} & \frac{2}{5} & \frac{5}{8} \\ \frac{3}{10} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$$

The total number of individuals in the migration process is 704*. After a long time, how many are in each location?*

Exercise 6.3.9 *You own a trailer rental company in a large city and you have four locations, one in the South East, one in the North East, one in the North West, and one in the South West. Denote these locations by SE,NE,NW, and SW respectively. Suppose that the following table is observed to take place.*

	SE	NE	NW	SW
SE	$\frac{1}{3}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{5}$
NE	$\frac{1}{3}$	$\frac{7}{10}$	$\frac{1}{5}$	$\frac{1}{10}$
NW	$\frac{2}{9}$	$\frac{1}{10}$	$\frac{3}{5}$	$\frac{1}{5}$
SW	$\frac{1}{9}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{2}$

In this table, the probability that a trailer starting at NE ends in NW is $1/10$, *the probability that a trailer starting at SW ends in NW is* $1/5$, *and so forth. Approximately how many will you have in each location after a long time if the total number of trailers is* 413*?*

Exercise 6.3.10 *You own a trailer rental company in a large city and you have four locations, one in the South East, one in the North East, one in the North West, and one in the South West. Denote these locations by SE,NE,NW, and SW respectively. Suppose that the following table is observed to take place.*

	SE	NE	NW	SW
SE	$\frac{1}{7}$	$\frac{1}{4}$	$\frac{1}{10}$	$\frac{1}{5}$
NE	$\frac{2}{7}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{10}$
NW	$\frac{1}{7}$	$\frac{1}{4}$	$\frac{3}{5}$	$\frac{1}{5}$
SW	$\frac{3}{7}$	$\frac{1}{4}$	$\frac{1}{10}$	$\frac{1}{2}$

In this table, the probability that a trailer starting at NE ends in NW is $1/10$, *the probability that a trailer starting at SW ends in NW is* $1/5$, *and so forth. Approximately how many will you have in each location after a long time if the total number of trailers is* 1469.

Exercise 6.3.11 *The following table describes the transition probabilities between the states rainy, partly cloudy and sunny. The symbol p.c. indicates partly cloudy. Thus if it starts off p.c. it ends up sunny the next day with probability* $\frac{1}{5}$. *If it starts off sunny, it ends up sunny the next day with probability* $\frac{2}{5}$ *and so forth.*

	rains	*sunny*	*p.c.*
rains	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{3}$
sunny	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{3}$
p.c.	$\frac{3}{5}$	$\frac{2}{5}$	$\frac{1}{3}$

Given this information, what are the probabilities that a given day is rainy, sunny, or partly cloudy?

Exercise 6.3.12 *The following table describes the transition probabilities between the states rainy, partly cloudy and sunny. The symbol p.c. indicates partly cloudy. Thus if it starts off p.c. it ends up sunny the next day with probability* $\frac{1}{10}$. *If it starts off sunny, it ends up sunny the next day with probability* $\frac{2}{5}$ *and so forth.*

	rains	*sunny*	*p.c.*
rains	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{3}$
sunny	$\frac{1}{10}$	$\frac{2}{5}$	$\frac{4}{9}$
p.c.	$\frac{7}{10}$	$\frac{2}{5}$	$\frac{2}{9}$

Given this information, what are the probabilities that a given day is rainy, sunny, or partly cloudy?

Exercise 6.3.13 *You own a trailer rental company in a large city and you have four locations, one in the South East, one in the North East, one in the North West, and one in the South West. Denote these locations by SE,NE,NW, and SW respectively. Suppose that the following table is observed to take place.*

	SE	NE	NW	SW
SE	$\frac{5}{11}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{5}$
NE	$\frac{1}{11}$	$\frac{7}{10}$	$\frac{1}{5}$	$\frac{1}{10}$
NW	$\frac{2}{11}$	$\frac{1}{10}$	$\frac{3}{5}$	$\frac{1}{5}$
SW	$\frac{3}{11}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{2}$

In this table, the probability that a trailer starting at NE ends in NW is $1/10$, *the probability that a trailer starting at SW ends in NW is* $1/5$, *and so forth. Approximately how many will you have in each location after a long time if the total number of trailers is* 407?

Exercise 6.3.14 *The University of Poohbah offers three degree programs, scouting education (SE), dance appreciation (DA), and engineering (E). It has been determined that the probabilities of transferring from one program to another are as in the following table.*

	SE	DA	E
SE	.8	.1	.3
DA	.1	.7	.5
E	.1	.2	.2

where the number indicates the probability of transferring from the top program to the program on the left. Thus the probability of going from DA to E is .2. Find the probability that a student is enrolled in the various programs.

Exercise 6.3.15 *In the city of Nabal, there are three political persuasions, republicans (R), democrats (D), and neither one (N). The following table shows the transition probabilities between the political parties, the top row being the initial political party and the side row being the political affiliation the following year.*

	R	D	N
R	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{2}{7}$
D	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{4}{7}$
N	$\frac{3}{5}$	$\frac{1}{2}$	$\frac{1}{7}$

Find the probabilities that a person will be identified with the various political persuasions. Which party will end up being most important?

Exercise 6.3.16 *The following table describes the transition probabilities between the states rainy, partly cloudy and sunny. The symbol p.c. indicates partly cloudy. Thus if it starts off p.c. it ends up sunny the next day with probability* $\frac{1}{5}$. *If it starts off sunny, it ends up sunny the next day with probability* $\frac{2}{7}$ *and so forth.*

	rains	*sunny*	*p.c.*
rains	$\frac{1}{5}$	$\frac{2}{7}$	$\frac{5}{9}$
sunny	$\frac{1}{5}$	$\frac{2}{7}$	$\frac{1}{3}$
p.c.	$\frac{3}{5}$	$\frac{3}{7}$	$\frac{1}{9}$

Given this information, what are the probabilities that a given day is rainy, sunny, or partly cloudy?

A. Some Prerequisite Topics

The topics presented in this section are important concepts in mathematics and therefore should be examined.

A.1 Sets and Set Notation

A set is a collection of things called elements. For example $\{1,2,3,8\}$ would be a set consisting of the elements 1,2,3, and 8. To indicate that 3 is an element of $\{1,2,3,8\}$, it is customary to write $3 \in \{1,2,3,8\}$. We can also indicate when an element is not in a set, by writing $9 \notin \{1,2,3,8\}$ which says that 9 is not an element of $\{1,2,3,8\}$. Sometimes a rule specifies a set. For example you could specify a set as all integers larger than 2. This would be written as $S = \{x \in \mathbb{Z} : x > 2\}$. This notation says: S is the set of all integers, x, such that $x > 2$.

Suppose A and B are sets with the property that every element of A is an element of B. Then we say that A is a subset of B. For example, $\{1,2,3,8\}$ is a subset of $\{1,2,3,4,5,8\}$. In symbols, we write $\{1,2,3,8\} \subseteq \{1,2,3,4,5,8\}$. It is sometimes said that "A is contained in B" or even "B contains A". The same statement about the two sets may also be written as $\{1,2,3,4,5,8\} \supseteq \{1,2,3,8\}$.

We can also talk about the *union* of two sets, which we write as $A \cup B$. This is the set consisting of everything which is an element of at least one of the sets, A or B. As an example of the union of two sets, consider $\{1,2,3,8\} \cup \{3,4,7,8\} = \{1,2,3,4,7,8\}$. This set is made up of the numbers which are in at least one of the two sets.

In general

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

Notice that an element which is in *both* A and B is also in the union, as well as elements which are in only one of A or B.

Another important set is the intersection of two sets A and B, written $A \cap B$. This set consists of everything which is in *both* of the sets. Thus $\{1,2,3,8\} \cap \{3,4,7,8\} = \{3,8\}$ because 3 and 8 are those elements the two sets have in common. In general,

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

If A and B are two sets, $A \setminus B$ denotes the set of things which are in A but not in B. Thus

$$A \setminus B = \{x \in A : x \notin B\}$$

For example, if $A = \{1,2,3,8\}$ and $B = \{3,4,7,8\}$, then $A \setminus B = \{1,2,3,8\} \setminus \{3,4,7,8\} = \{1,2\}$.

A special set which is very important in mathematics is the empty set denoted by $\emptyset$. The empty set, $\emptyset$, is defined as the set which has no elements in it. It follows that the empty set is a subset of every set. This is true because if it were not so, there would have to exist a set A, such that $\emptyset$ has something in it which is not in A. However, $\emptyset$ has nothing in it and so it must be that $\emptyset \subseteq A$.

We can also use brackets to denote sets which are intervals of numbers. Let a and b be real numbers. Then

- $[a,b] = \{x \in \mathbb{R} : a \leq x \leq b\}$
- $[a,b) = \{x \in \mathbb{R} : a \leq x < b\}$
- $(a,b) = \{x \in \mathbb{R} : a < x < b\}$
- $(a,b] = \{x \in \mathbb{R} : a < x \leq b\}$
- $[a,\infty) = \{x \in \mathbb{R} : x \geq a\}$
- $(-\infty,a] = \{x \in \mathbb{R} : x \leq a\}$

These sorts of sets of real numbers are called intervals. The two points a and b are called endpoints, or bounds, of the interval. In particular, a is the *lower bound* while b is the *upper bound* of the above intervals, where applicable. Other intervals such as $(-\infty, b)$ are defined by analogy to what was just explained. In general, the curved parenthesis, (, indicates the end point is not included in the interval, while the square parenthesis, [, indicates this end point is included. The reason that there will always be a curved parenthesis next to ∞ or $-\infty$ is that these are not real numbers and cannot be included in the interval in the way a real number can.

To illustrate the use of this notation relative to intervals consider three examples of inequalities. Their solutions will be written in the interval notation just described.

Example A.1: Solving an Inequality

Solve the inequality $2x + 4 \leq x - 8$.

Solution. We need to find x such that $2x + 4 \leq x - 8$. Solving for x, we see that $x \leq -12$ is the answer. This is written in terms of an interval as $(-\infty, -12]$. ☐

Consider the following example.

Example A.2: Solving an Inequality

Solve the inequality $(x + 1)(2x - 3) \geq 0$.

Solution. We need to find x such that $(x + 1)(2x - 3) \geq 0$. The solution is given by $x \leq -1$ or $x \geq \frac{3}{2}$. Therefore, x which fit into either of these intervals gives a solution. In terms of set notation this is denoted by $(-\infty, -1] \cup [\frac{3}{2}, \infty)$. ☐

Consider one last example.

Example A.3: Solving an Inequality

Solve the inequality $x(x+2) \geq -4$.

Solution. This inequality is true for any value of x where x is a real number. We can write the solution as $\mathbb{R}$ or $(-\infty, \infty)$. ☐

In the next section, we examine another important mathematical concept.

A.2 Well Ordering and Induction

We begin this section with some important notation. Summation notation, written $\sum_{i=1}^{j} i$, represents a sum. Here, i is called the index of the sum, and we add iterations until $i = j$. For example,

$$\sum_{i=1}^{j} i = 1 + 2 + \cdots + j$$

Another example:

$$a_{11} + a_{12} + a_{13} = \sum_{i=1}^{3} a_{1i}$$

The following notation is a specific use of summation notation.

Notation A.4: Summation Notation

Let a_{ij} be real numbers, and suppose $1 \leq i \leq r$ while $1 \leq j \leq s$. These numbers can be listed in a rectangular array as given by

$$\begin{matrix} a_{11} & a_{12} & \cdots & a_{1s} \\ a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rs} \end{matrix}$$

Then $\sum_{j=1}^{s} \sum_{i=1}^{r} a_{ij}$ means to first sum the numbers in each column (using i as the index) and then to add the sums which result (using j as the index). Similarly, $\sum_{i=1}^{r} \sum_{j=1}^{s} a_{ij}$ means to sum the vectors in each row (using j as the index) and then to add the sums which result (using i as the index).

Notice that since addition is commutative, $\sum_{j=1}^{s} \sum_{i=1}^{r} a_{ij} = \sum_{i=1}^{r} \sum_{j=1}^{s} a_{ij}$.

We now consider the main concept of this section. Mathematical induction and well ordering are two extremely important principles in math. They are often used to prove significant things which would be hard to prove otherwise.

Definition A.5: Well Ordered

A set is well ordered if every nonempty subset S, contains a smallest element z having the property that $z \leq x$ for all $x \in S$.

In particular, the set of natural numbers defined as

$$\mathbb{N} = \{1, 2, \cdots\}$$

is well ordered.

Consider the following proposition.

Proposition A.6: Well Ordered Sets

Any set of integers larger than a given number is well ordered.

This proposition claims that if a set has a lower bound which is a real number, then this set is well ordered.

Further, this proposition implies the principle of mathematical induction. The symbol $\mathbb{Z}$ denotes the set of all integers. Note that if a is an integer, then there are no integers between a and $a+1$.

Theorem A.7: Mathematical Induction

A set $S \subseteq \mathbb{Z}$, having the property that $a \in S$ and $n+1 \in S$ whenever $n \in S$, contains all integers $x \in \mathbb{Z}$ such that $x \geq a$.

Proof. Let T consist of all integers larger than or equal to a which are not in S. The theorem will be proved if $T = \emptyset$. If $T \neq \emptyset$ then by the well ordering principle, there would have to exist a smallest element of T, denoted as b. It must be the case that $b > a$ since by definition, $a \notin T$. Thus $b \geq a+1$, and so $b-1 \geq a$ and $b-1 \notin S$ because if $b-1 \in S$, then $b-1+1 = b \in S$ by the assumed property of S. Therefore, $b-1 \in T$ which contradicts the choice of b as the smallest element of T. ($b-1$ is smaller.) Since a contradiction is obtained by assuming $T \neq \emptyset$, it must be the case that $T = \emptyset$ and this says that every integer at least as large as a is also in S. $\square$

Mathematical induction is a very useful device for proving theorems about the integers. The procedure is as follows.

Procedure A.8: Proof by Mathematical Induction

Suppose S_n is a statement which is a function of the number n, for $n = 1, 2, \cdots$, and we wish to show that S_n is true for all $n \geq 1$. To do so using mathematical induction, use the following steps.

1. ***Base Case:*** *Show S_1 is true.*
2. *Assume S_n is true for some n, which is the **induction hypothesis**. Then, using this assumption, show that S_{n+1} is true.*

Proving these two steps shows that S_n is true for all $n = 1, 2, \cdots$.

We can use this procedure to solve the following examples.

Example A.9: Proving by Induction

Prove by induction that $\sum_{k=1}^{n} k^2 = \dfrac{n(n+1)(2n+1)}{6}$.

Solution. By Procedure A.8, we first need to show that this statement is true for $n = 1$. When $n = 1$, the statement says that

$$\begin{aligned} \sum_{k=1}^{1} k^2 &= \frac{1(1+1)(2(1)+1)}{6} \\ &= \frac{6}{6} \\ &= 1 \end{aligned}$$

The sum on the left hand side also equals 1, so this equation is true for $n = 1$.

Now suppose this formula is valid for some $n \geq 1$ where n is an integer. Hence, the following equation is true.

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} \tag{1.1}$$

We want to show that this is true for $n + 1$.

Suppose we add $(n+1)^2$ to both sides of equation 1.1.

$$\begin{aligned} \sum_{k=1}^{n+1} k^2 &= \sum_{k=1}^{n} k^2 + (n+1)^2 \\ &= \frac{n(n+1)(2n+1)}{6} + (n+1)^2 \end{aligned}$$

The step going from the first to the second line is based on the assumption that the formula is true for n. Now simplify the expression in the second line,

$$\frac{n(n+1)(2n+1)}{6} + (n+1)^2$$

This equals

$$(n+1)\left(\frac{n(2n+1)}{6}+(n+1)\right)$$

and

$$\frac{n(2n+1)}{6}+(n+1)=\frac{6(n+1)+2n^2+n}{6}=\frac{(n+2)(2n+3)}{6}$$

Therefore,

$$\sum_{k=1}^{n+1} k^2=\frac{(n+1)(n+2)(2n+3)}{6}=\frac{(n+1)((n+1)+1)(2(n+1)+1)}{6}$$

showing the formula holds for $n+1$ whenever it holds for n. This proves the formula by mathematical induction. In other words, this formula is true for all $n=1,2,\cdots$. ☐

Consider another example.

Example A.10: Proving an Inequality by Induction

Show that for all $n \in \mathbb{N}$, $\frac{1}{2}\cdot\frac{3}{4}\cdots\frac{2n-1}{2n}<\frac{1}{\sqrt{2n+1}}$.

Solution. Again we will use the procedure given in Procedure A.8 to prove that this statement is true for all n. Suppose $n=1$. Then the statement says

$$\frac{1}{2}<\frac{1}{\sqrt{3}}$$

which is true.

Suppose then that the inequality holds for n. In other words,

$$\frac{1}{2}\cdot\frac{3}{4}\cdots\frac{2n-1}{2n}<\frac{1}{\sqrt{2n+1}}$$

is true.

Now multiply both sides of this inequality by $\frac{2n+1}{2n+2}$. This yields

$$\frac{1}{2}\cdot\frac{3}{4}\cdots\frac{2n-1}{2n}\cdot\frac{2n+1}{2n+2}<\frac{1}{\sqrt{2n+1}}\frac{2n+1}{2n+2}=\frac{\sqrt{2n+1}}{2n+2}$$

The theorem will be proved if this last expression is less than $\frac{1}{\sqrt{2n+3}}$. This happens if and only if

$$\left(\frac{1}{\sqrt{2n+3}}\right)^2=\frac{1}{2n+3}>\frac{2n+1}{(2n+2)^2}$$

which occurs if and only if $(2n+2)^2>(2n+3)(2n+1)$ and this is clearly true which may be seen from expanding both sides. This proves the inequality. ☐

Let's review the process just used. If S is the set of integers at least as large as 1 for which the formula holds, the first step was to show $1 \in S$ and then that whenever $n \in S$, it follows $n + 1 \in S$. Therefore, by the principle of mathematical induction, S contains $[1, \infty) \cap \mathbb{Z}$, all positive integers. In doing an inductive proof of this sort, the set S is normally not mentioned. One just verifies the steps above.

B. Selected Exercise Answers

1.1.1 $\begin{array}{c} x+3y=1 \\ 4x-y=3 \end{array}$, Solution is: $\left[x=\frac{10}{13}, y=\frac{1}{13}\right]$.

1.1.2 $\begin{array}{c} 3x+y=3 \\ x+2y=1 \end{array}$, Solution is: $[x=1, y=0]$

1.2.1 $\begin{array}{c} x+3y=1 \\ 4x-y=3 \end{array}$, Solution is: $\left[x=\frac{10}{13}, y=\frac{1}{13}\right]$

1.2.2 $\begin{array}{c} x+3y=1 \\ 4x-y=3 \end{array}$, Solution is: $\left[x=\frac{10}{13}, y=\frac{1}{13}\right]$

1.2.3 $\begin{array}{c} x+2y=1 \\ 2x-y=1 \\ 4x+3y=3 \end{array}$, Solution is: $\left[x=\frac{3}{5}, y=\frac{1}{5}\right]$

1.2.4 No solution exists. You can see this by writing the augmented matrix and doing row operations. $\left[\begin{array}{rrrr} 1 & 1 & -3 & 2 \\ 2 & 1 & 1 & 1 \\ 3 & 2 & -2 & 0 \end{array}\right]$, row echelon form: $\left[\begin{array}{rrrr} 1 & 0 & 4 & 0 \\ 0 & 1 & -7 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$. Thus one of the equations says $0=1$ in an equivalent system of equations.

1.2.5 $\begin{array}{c} 4g-I=150 \\ 4I-17g=-660 \\ 4g+s=290 \\ g+I+s-b=0 \end{array}$, Solution is : $\{g=60, I=90, b=200, s=50\}$

1.2.6 The solution exists but is not unique.

1.2.7 A solution exists and is unique.

1.2.9 There might be a solution. If so, there are infinitely many.

1.2.10 No. Consider $x+y+z=2$ and $x+y+z=1$.

1.2.11 These can have a solution. For example, $x+y=1, 2x+2y=2, 3x+3y=3$ even has an infinite set of solutions.

1.2.12 $h=4$

1.2.13 Any h will work.

1.2.14 Any h will work.

1.2.15 If $h \neq 2$ there will be a unique solution for any k. If $h = 2$ and $k \neq 4$, there are no solutions. If $h = 2$ and $k = 4$, then there are infinitely many solutions.

1.2.16 If $h \neq 4$, then there is exactly one solution. If $h = 4$ and $k \neq 4$, then there are no solutions. If $h = 4$ and $k = 4$, then there are infinitely many solutions.

1.2.17 There is no solution. The system is inconsistent. You can see this from the augmented matrix. $\begin{bmatrix} 1 & 2 & 1 & -1 & 2 \\ 1 & -1 & 1 & 1 & 1 \\ 2 & 1 & -1 & 0 & 1 \\ 4 & 2 & 1 & 0 & 5 \end{bmatrix}$, reduced row-echelon form: $\begin{bmatrix} 1 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 1 & 0 & -\frac{2}{3} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$.

1.2.18 Solution is: $\left[w = \frac{3}{2}y - 1, x = \frac{2}{3} - \frac{1}{2}y, z = \frac{1}{3}\right]$

1.2.19
1. This one is not.
2. This one is.
3. This one is.

1.2.28 The reduced row-echelon form is $\left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & -\frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 0 \end{array}\right]$. Therefore, the solution is of the form $z = t, y = \frac{3}{4} + t\left(\frac{1}{4}\right), x = \frac{1}{2} - \frac{1}{2}t$ where $t \in \mathbb{R}$.

1.2.29 The reduced row-echelon form is $\left[\begin{array}{ccc|c} 1 & 0 & 4 & 2 \\ 0 & 1 & -4 & -1 \end{array}\right]$ and so the solution is $z = t, y = 4t, x = 2 - 4t$.

1.2.30 The reduced row-echelon form is $\left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 9 & 3 \\ 0 & 1 & 0 & 0 & -4 & 0 \\ 0 & 0 & 1 & 0 & -7 & -1 \\ 0 & 0 & 0 & 1 & 6 & 1 \end{array}\right]$ and so $x_5 = t, x_4 = 1 - 6t, x_3 = -1 + 7t, x_2 = 4t, x_1 = 3 - 9t$.

1.2.31 The reduced row-echelon form is $\left[\begin{array}{ccccc|c} 1 & 0 & 2 & 0 & -\frac{1}{2} & \frac{5}{2} \\ 0 & 1 & 0 & 0 & \frac{1}{2} & \frac{3}{2} \\ 0 & 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$. Therefore, let $x_5 = t, x_3 = s$. Then the other variables are given by $x_4 = -\frac{1}{2} - \frac{3}{2}t, x_2 = \frac{3}{2} - t\frac{1}{2},, x_1 = \frac{5}{2} + \frac{1}{2}t - 2s$.

1.2.32 Solution is: $[x = 1 - 2t, z = 1, y = t]$

1.2.33 Solution is: $[x = 2 - 4t, y = -8t, z = t]$

1.2.34 Solution is: $[x = -1, y = 2, z = -1]$

1.2.35 Solution is: $[x = 2, y = 4, z = 5]$

1.2.36 Solution is: $[x = 1, y = 2, z = -5]$

1.2.37 Solution is: $[x = -1, y = -5, z = 4]$

1.2.38 Solution is: $[x = 2t + 1, y = 4t, z = t]$

1.2.39 Solution is: $[x = 1, y = 5, z = 3]$

1.2.40 Solution is: $[x = 4, y = -4, z = -2]$

1.2.41 No. Consider $x + y + z = 2$ and $x + y + z = 1$.

1.2.42 No. This would lead to $0 = 1$.

1.2.43 Yes. It has a unique solution.

1.2.44 The last column must not be a pivot column. The remaining columns must each be pivot columns.

1.2.45 You need $\begin{array}{c} \frac{1}{4}(20 + 30 + w + x) - y = 0 \\ \frac{1}{4}(y + 30 + 0 + z) - w = 0 \\ \frac{1}{4}(20 + y + z + 10) - x = 0 \\ \frac{1}{4}(x + w + 0 + 10) - z = 0 \end{array}$, Solution is: $[w = 15, x = 15, y = 20, z = 10]$.

1.2.46 The other two equations are

$$\begin{array}{rcl} 6I_3 - 6I_4 + I_3 + I_3 + 5I_3 - 5I_2 & = & -20 \\ 2I_4 + 3I_4 + 6I_4 - 6I_3 + I_4 - I_1 & = & 0 \end{array}$$

Then the system is

$$\begin{array}{c} 2I_2 - 2I_1 + 5I_2 - 5I_3 + 3I_2 = 5 \\ 4I_1 + I_1 - I_4 + 2I_1 - 2I_2 = -10 \\ 6I_3 - 6I_4 + I_3 + I_3 + 5I_3 - 5I_2 = -20 \\ 2I_4 + 3I_4 + 6I_4 - 6I_3 + I_4 - I_1 = 0 \end{array}$$

The solution is:

$$I_1 = -2.0107, I_2 = -1.2699, I_3 = -2.7355, I_4 = -1.5353$$

1.2.47 You have

$$\begin{aligned} 2I_1+5I_1+3I_1-5I_2 &= -10 \\ I_2+3I_2+7I_2+5I_2-5I_1 &= -12 \\ 2I_3+4I_3+4I_3+I_3-I_2 &= 0 \end{aligned}$$

Simplifying this yields

$$\begin{aligned} 10I_1-5I_2 &= -10 \\ 16I_2-5I_1 &= -12 \\ 11I_3-I_2 &= 0 \end{aligned}$$

Then you just find the solution.

$$I_1=-\frac{44}{27}, I_2=-\frac{34}{27}, I_3=-\frac{34}{297}$$

Thus all currents flow in the clockwise direction.

1.2.59 It is because you cannot have more than $\min(m,n)$ nonzero rows in the reduced row-echelon form. Recall that the number of pivot columns is the same as the number of nonzero rows from the description of this reduced row-echelon form.

1.2.60

1. This says B is in the span of four of the columns. Thus the columns are not independent. Infinite solution set.
2. This surely can't happen. If you add in another column, the rank does not get smaller.
3. This says B is in the span of the columns and the columns must be independent. You can't have the rank equal 4 if you only have two columns.
4. This says B is not in the span of the columns. In this case, there is no solution to the system of equations represented by the augmented matrix.
5. In this case, there is a unique solution since the columns of A are independent.

1.2.61 These are not legitimate row operations. They do not preserve the solution set of the system.

2.1.3 To get $-A$, just replace every entry of A with its additive inverse. The 0 matrix is the one which has all zeros in it.

2.1.5 Suppose B also works. Then

$$-A=-A+(A+B)=(-A+A)+B=0+B=B$$

2.1.6 Suppose $0'$ also works. Then $0'=0'+0=0$.

2.1.7 $0A=(0+0)A=0A+0A$. Now add $-(0A)$ to both sides. Then $0=0A$.

2.1.8 $A+(-1)\,A=(1+(-1))\,A=0A=0$. Therefore, from the uniqueness of the additive inverse proved in the above Problem 2.1.5, it follows that $-A=(-1)\,A$.

2.1.9 1. $\begin{bmatrix} -3 & -6 & -9 \\ -6 & -3 & -21 \end{bmatrix}$

2. $\begin{bmatrix} 8 & -5 & 3 \\ -11 & 5 & -4 \end{bmatrix}$

3. Not possible

4. $\begin{bmatrix} -3 & 3 & 4 \\ 6 & -1 & 7 \end{bmatrix}$

5. Not possible

6. Not possible

2.1.10 1. $\begin{bmatrix} -3 & -6 \\ -9 & -6 \\ -3 & 3 \end{bmatrix}$

2. Not possible.

3. $\begin{bmatrix} 11 & 2 \\ 13 & 6 \\ -4 & 2 \end{bmatrix}$

4. Not possible.

5. $\begin{bmatrix} 7 \\ 9 \\ -2 \end{bmatrix}$

6. Not possible.

7. Not possible.

8. $\begin{bmatrix} 2 \\ -5 \end{bmatrix}$

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

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

3. Not possible

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

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

2.1.12

$$\begin{aligned} \begin{bmatrix} -1 & -1 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} x & y \\ z & w \end{bmatrix} &= \begin{bmatrix} -x-z & -w-y \\ 3x+3z & 3w+3y \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

Solution is: $w = -y, x = -z$ so the matrices are of the form $\begin{bmatrix} x & y \\ -x & -y \end{bmatrix}$.

2.1.13 $X^T Y = \begin{bmatrix} 0 & -1 & -2 \\ 0 & -1 & -2 \\ 0 & 1 & 2 \end{bmatrix}, XY^T = 1$

2.1.14

$$\begin{aligned} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & k \end{bmatrix} &= \begin{bmatrix} 7 & 2k+2 \\ 15 & 4k+6 \end{bmatrix} \\ \begin{bmatrix} 1 & 2 \\ 3 & k \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} &= \begin{bmatrix} 7 & 10 \\ 3k+3 & 4k+6 \end{bmatrix} \end{aligned}$$

Thus you must have $\begin{array}{c} 3k+3 = 15 \\ 2k+2 = 10 \end{array}$, Solution is: $[k = 4]$

2.1.15

$$\begin{aligned} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & k \end{bmatrix} &= \begin{bmatrix} 3 & 2k+2 \\ 7 & 4k+6 \end{bmatrix} \\ \begin{bmatrix} 1 & 2 \\ 1 & k \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} &= \begin{bmatrix} 7 & 10 \\ 3k+1 & 4k+2 \end{bmatrix} \end{aligned}$$

However, $7 \neq 3$ and so there is no possible choice of k which will make these matrices commute.

2.1.16 Let $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, C = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$.

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

2.1.18 Let $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

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

2.1.20 Let $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

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

2.1.21 $A = \begin{bmatrix} 1 & -1 & 2 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 0 & 3 & 0 \\ 1 & 3 & 0 & 3 \end{bmatrix}$

2.1.22 $A = \begin{bmatrix} 1 & 3 & 2 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 0 & 6 & 0 \\ 1 & 3 & 0 & 1 \end{bmatrix}$

2.1.23 $A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 3 \end{bmatrix}$

2.1.26

1. Not necessarily true.
2. Not necessarily true.
3. Not necessarily true.
4. Necessarily true.
5. Necessarily true.
6. Not necessarily true.
7. Not necessarily true.

2.1.27

1. $\begin{bmatrix} -3 & -9 & -3 \\ -6 & -6 & 3 \end{bmatrix}$
2. $\begin{bmatrix} 5 & -18 & 5 \\ -11 & 4 & 4 \end{bmatrix}$

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

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

5. $\begin{bmatrix} 13 & -16 & 1 \\ -16 & 29 & -8 \\ 1 & -8 & 5 \end{bmatrix}$

6. $\begin{bmatrix} 5 & 7 & -1 \\ 5 & 15 & 5 \end{bmatrix}$

7. Not possible.

2.1.28 Show that $\frac{1}{2}\left(A^T + A\right)$ is symmetric and then consider using this as one of the matrices. $A = \frac{A+A^T}{2} + \frac{A-A^T}{2}$.

2.1.29 If A is symmetric then $A = -A^T$. It follows that $a_{ii} = -a_{ii}$ and so each $a_{ii} = 0$.

2.1.31 $(I_m A)_{ij} \equiv \sum_j \delta_{ik} A_{kj} = A_{ij}$

2.1.32 Yes $B = C$. Multiply $AB = AC$ on the left by A^{-1}.

2.1.34 $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

2.1.35 $\begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{3}{7} & -\frac{1}{7} \\ \frac{1}{7} & \frac{2}{7} \end{bmatrix}$

2.1.36 $\begin{bmatrix} 0 & 1 \\ 5 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} -\frac{3}{5} & \frac{1}{5} \\ 1 & 0 \end{bmatrix}$

2.1.37 $\begin{bmatrix} 2 & 1 \\ 3 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & \frac{1}{3} \\ 1 & -\frac{2}{3} \end{bmatrix}$

2.1.38 $\begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}^{-1}$ does not exist. The reduced row-echelon form of this matrix is $\begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 0 \end{bmatrix}$

2.1.39 $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}$

2.1.40 $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 1 & 0 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} -2 & 4 & -5 \\ 0 & 1 & -2 \\ 1 & -2 & 3 \end{bmatrix}$

2.1.41 $\begin{bmatrix} 1 & 0 & 3 \\ 2 & 3 & 4 \\ 1 & 0 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} -2 & 0 & 3 \\ 0 & \frac{1}{3} & -\frac{2}{3} \\ 1 & 0 & -1 \end{bmatrix}$

2.1.42 The reduced row-echelon form is $\begin{bmatrix} 1 & 0 & \frac{5}{3} \\ 0 & 1 & \frac{2}{3} \\ 0 & 0 & 0 \end{bmatrix}$. There is no inverse.

2.1.43 $\begin{bmatrix} 1 & 2 & 0 & 2 \\ 1 & 1 & 2 & 0 \\ 2 & 1 & -3 & 2 \\ 1 & 2 & 1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} -1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 3 & \frac{1}{2} & -\frac{1}{2} & -\frac{5}{2} \\ -1 & 0 & 0 & 1 \\ -2 & -\frac{3}{4} & \frac{1}{4} & \frac{9}{4} \end{bmatrix}$

2.1.45 1. $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ -\frac{2}{3} \\ 0 \end{bmatrix}$

2. $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -12 \\ 1 \\ 5 \end{bmatrix}$

$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3c - 2a \\ \frac{1}{3}b - \frac{2}{3}c \\ a - c \end{bmatrix}$

2.1.46 Multiply both sides of $AX = B$ on the left by A^{-1}.

2.1.47 Multiply on both sides on the left by A^{-1}. Thus

$$0 = A^{-1}0 = A^{-1}(AX) = \left(A^{-1}A\right)X = IX = X$$

2.1.48 $A^{-1} = A^{-1}I = A^{-1}(AB) = (A^{-1}A)B = IB = B$.

2.1.49 You need to show that $(A^{-1})^T$ acts like the inverse of A^T because from uniqueness in the above problem, this will imply it is the inverse. From properties of the transpose,

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

Hence $(A^{-1})^T = \left(A^T\right)^{-1}$ and this last matrix exists.

2.1.50 $(AB)B^{-1}A^{-1} = A(BB^{-1})A^{-1} = AA^{-1} = I$ $B^{-1}A^{-1}(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I$

2.1.51 The proof of this exercise follows from the previous one.

2.1.52 $A^2 \left(A^{-1}\right)^2 = AAA^{-1}A^{-1} = AIA^{-1} = AA^{-1} = I \left(A^{-1}\right)^2 A^2 = A^{-1}A^{-1}AA = A^{-1}IA = A^{-1}A = I$

2.1.53 $A^{-1}A = AA^{-1} = I$ and so by uniqueness, $\left(A^{-1}\right)^{-1} = A$.

3.1.3

1. The answer is 31.
2. The answer is 375.
3. The answer is -2.

3.1.4

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix} = 6$$

3.1.5

$$\begin{vmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{vmatrix} = 2$$

3.1.6

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix} = 6$$

3.1.7

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

3.1.9 It does not change the determinant. This was just taking the transpose.

3.1.10 In this case two rows were switched and so the resulting determinant is -1 times the first.

3.1.11 The determinant is unchanged. It was just the first row added to the second.

3.1.12 The second row was multiplied by 2 so the determinant of the result is 2 times the original determinant.

3.1.13 In this case the two columns were switched so the determinant of the second is -1 times the determinant of the first.

3.1.14 If the determinant is nonzero, then it will remain nonzero with row operations applied to the matrix. However, by assumption, you can obtain a row of zeros by doing row operations. Thus the determinant must have been zero after all.

3.1.15 $\det(aA) = \det(aIA) = \det(aI)\det(A) = a^n \det(A)$. The matrix which has a down the main diagonal has determinant equal to a^n.

3.1.16

$$\det\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} -1 & 2 \\ -5 & 6 \end{bmatrix}\right) = -8$$

$$\det\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\det\begin{bmatrix} -1 & 2 \\ -5 & 6 \end{bmatrix} = -2 \times 4 = -8$$

3.1.17 This is not true at all. Consider $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$.

3.1.18 It must be 0 because $0 = \det(0) = \det\left(A^k\right) = (\det(A))^k$.

3.1.19 You would need $\det\left(AA^T\right) = \det(A)\det\left(A^T\right) = \det(A)^2 = 1$ and so $\det(A) = 1$, or -1.

3.1.20 $\det(A) = \det(S^{-1}BS) = \det(S^{-1})\det(B)\det(S) = \det(B)\det(S^{-1}S) = \det(B)$.

3.1.21

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

3.1.22

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

3.1.23

$$\begin{vmatrix} 2 & 1 & 3 \\ 2 & 4 & 2 \\ 1 & 4 & -5 \end{vmatrix} = -32$$

3.1.24 One can row reduce this using only row operation 3 to

$$\begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & -5 & -5 & -3 \\ 0 & 0 & 2 & \frac{9}{5} \\ 0 & 0 & 0 & -\frac{63}{10} \end{bmatrix}$$

and therefore, the determinant is -63.

$$\begin{vmatrix} 1 & 2 & 1 & 2 \\ 3 & 1 & -2 & 3 \\ -1 & 0 & 3 & 1 \\ 2 & 3 & 2 & -2 \end{vmatrix} = 63$$

3.1.25 One can row reduce this using only row operation 3 to

$$\begin{bmatrix} 1 & 4 & 1 & 2 \\ 0 & -10 & -5 & -3 \\ 0 & 0 & 2 & \frac{19}{5} \\ 0 & 0 & 0 & -\frac{211}{20} \end{bmatrix}$$

Thus the determinant is given by

$$\begin{vmatrix} 1 & 4 & 1 & 2 \\ 3 & 2 & -2 & 3 \\ -1 & 0 & 3 & 3 \\ 2 & 1 & 2 & -2 \end{vmatrix} = 211$$

3.2.1 $\det \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 3 & 1 & 0 \end{bmatrix} = -13$ and so it has an inverse. This inverse is

$$\frac{1}{-13}\begin{bmatrix} \begin{vmatrix} 2 & 1 \\ 1 & 0 \end{vmatrix} & -\begin{vmatrix} 0 & 1 \\ 3 & 0 \end{vmatrix} & \begin{vmatrix} 0 & 2 \\ 3 & 1 \end{vmatrix} \\ -\begin{vmatrix} 2 & 3 \\ 1 & 0 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 3 & 0 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} \\ \begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 0 & 1 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 0 & 2 \end{vmatrix} \end{bmatrix}^T = \frac{1}{-13}\begin{bmatrix} -1 & 3 & -6 \\ 3 & -9 & 5 \\ -4 & -1 & 2 \end{bmatrix}^T$$

$$= \begin{bmatrix} \frac{1}{13} & -\frac{3}{13} & \frac{4}{13} \\ -\frac{3}{13} & \frac{9}{13} & \frac{1}{13} \\ \frac{6}{13} & -\frac{5}{13} & -\frac{2}{13} \end{bmatrix}$$

3.2.2 $\det \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 1 \\ 3 & 1 & 1 \end{bmatrix} = 7$ so it has an inverse. This inverse is $\frac{1}{7}\begin{bmatrix} 1 & 3 & -6 \\ -2 & 1 & 5 \\ 2 & -1 & 2 \end{bmatrix}^T =$
$\begin{bmatrix} \frac{1}{7} & -\frac{2}{7} & \frac{2}{7} \\ \frac{3}{7} & \frac{1}{7} & -\frac{1}{7} \\ -\frac{6}{7} & \frac{5}{7} & \frac{2}{7} \end{bmatrix}$

3.2.3

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

so it has an inverse which is

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

3.2.5

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

and so it has an inverse. The inverse turns out to equal

$$\begin{bmatrix} -\frac{1}{2} & \frac{3}{2} & 0 \\ \frac{3}{2} & -\frac{9}{2} & 1 \\ \frac{1}{2} & -\frac{1}{2} & 0 \end{bmatrix}$$

3.2.6

1. $\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} = 1$
2. $\begin{vmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 4 & 1 & 1 \end{vmatrix} = -15$
3. $\begin{vmatrix} 1 & 2 & 1 \\ 2 & 3 & 0 \\ 0 & 1 & 2 \end{vmatrix} = 0$

3.2.7 No. It has a nonzero determinant for all t

3.2.8

$$\det\begin{bmatrix} 1 & t & t^2 \\ 0 & 1 & 2t \\ t & 0 & 2 \end{bmatrix} = t^3 + 2$$

and so it has no inverse when $t = -\sqrt[3]{2}$

3.2.9

$$\det\begin{bmatrix} e^t & \cosh t & \sinh t \\ e^t & \sinh t & \cosh t \\ e^t & \cosh t & \sinh t \end{bmatrix} = 0$$

and so this matrix fails to have a nonzero determinant at any value of t.

3.2.10

$$\det\begin{bmatrix} e^t & e^{-t}\cos t & e^{-t}\sin t \\ e^t & -e^{-t}\cos t - e^{-t}\sin t & -e^{-t}\sin t + e^{-t}\cos t \\ e^t & 2e^{-t}\sin t & -2e^{-t}\cos t \end{bmatrix} = 5e^{-t} \neq 0$$

and so this matrix is always invertible.

3.2.11 If $\det(A) \neq 0$, then A^{-1} exists and so you could multiply on both sides on the left by A^{-1} and obtain that $X = 0$.

3.2.12 You have $1 = \det(A)\det(B)$. Hence both A and B have inverses. Letting X be given,

$$A(BA - I)X = (AB)AX - AX = AX - AX = 0$$

and so it follows from the above problem that $(BA - I)X = 0$. Since X is arbitrary, it follows that $BA = I$.

3.2.13

$$\det\begin{bmatrix} e^t & 0 & 0 \\ 0 & e^t\cos t & e^t\sin t \\ 0 & e^t\cos t - e^t\sin t & e^t\cos t + e^t\sin t \end{bmatrix} = e^{3t}.$$

Hence the inverse is

$$e^{-3t}\begin{bmatrix} e^{2t} & 0 & 0 \\ 0 & e^{2t}\cos t + e^{2t}\sin t & -\left(e^{2t}\cos t - e^{2t}\sin\right)t \\ 0 & -e^{2t}\sin t & e^{2t}\cos(t) \end{bmatrix}^T$$

$$= \begin{bmatrix} e^{-t} & 0 & 0 \\ 0 & e^{-t}(\cos t + \sin t) & -(\sin t)e^{-t} \\ 0 & -e^{-t}(\cos t - \sin t) & (\cos t)e^{-t} \end{bmatrix}$$

3.2.14

$$\begin{bmatrix} e^t & \cos t & \sin t \\ e^t & -\sin t & \cos t \\ e^t & -\cos t & -\sin t \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \frac{1}{2}e^{-t} & 0 & \frac{1}{2}e^{-t} \\ \frac{1}{2}\cos t + \frac{1}{2}\sin t & -\sin t & \frac{1}{2}\sin t - \frac{1}{2}\cos t \\ \frac{1}{2}\sin t - \frac{1}{2}\cos t & \cos t & -\frac{1}{2}\cos t - \frac{1}{2}\sin t \end{bmatrix}$$

3.2.15 The given condition is what it takes for the determinant to be non zero. Recall that the determinant of an upper triangular matrix is just the product of the entries on the main diagonal.

3.2.16 This follows because $\det(ABC) = \det(A)\det(B)\det(C)$ and if this product is nonzero, then each determinant in the product is nonzero and so each of these matrices is invertible.

3.2.17 False.

3.2.18 Solution is: $[x = 1, y = 0]$

3.2.19 Solution is: $[x = 1, y = 0, z = 0]$. For example,

$$y = \frac{\begin{vmatrix} 1 & 1 & 1 \\ 2 & 2 & -1 \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 2 & -1 & -1 \\ 1 & 0 & 1 \end{vmatrix}} = 0$$

4.2.1 $\begin{bmatrix} -55 \\ 13 \\ -21 \\ 39 \end{bmatrix}$

4.2.3

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

4.2.4 The system

$$\begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} = a_1\begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix} + a_2\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

has no solution.

4.7.1 $\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \bullet \begin{bmatrix} 2 \\ 0 \\ 1 \\ 3 \end{bmatrix} = 17$

4.7.2 This formula says that $\vec{u} \bullet \vec{v} = \|\vec{u}\|\|\vec{v}\| \cos\theta$ where θ is the included angle between the two vectors. Thus

$$\|\vec{u} \bullet \vec{v}\| = \|\vec{u}\|\|\vec{v}\|\|\cos\theta\| \leq \|\vec{u}\|\|\vec{v}\|$$

and equality holds if and only if $\theta = 0$ or π. This means that the two vectors either point in the same direction or opposite directions. Hence one is a multiple of the other.

4.7.3 This follows from the Cauchy Schwarz inequality and the proof of Theorem 4.29 which only used the properties of the dot product. Since this new product has the same properties the Cauchy Schwarz inequality holds for it as well.

4.7.6 $A\vec{x} \bullet \vec{y} = \sum_k (A\vec{x})_k y_k = \sum_k \sum_i A_{ki} x_i y_k = \sum_i \sum_k A^T_{ik} x_i y_k = \vec{x} \bullet A^T \vec{y}$

4.7.7

$$\begin{aligned} AB\vec{x} \bullet \vec{y} &= B\vec{x} \bullet A^T \vec{y} \\ &= \vec{x} \bullet B^T A^T \vec{y} \\ &= \vec{x} \bullet (AB)^T \vec{y} \end{aligned}$$

Since this is true for all $\vec{x}$, it follows that, in particular, it holds for

$$\vec{x} = B^T A^T \vec{y} - (AB)^T \vec{y}$$

and so from the axioms of the dot product,

$$\left(B^T A^T \vec{y} - (AB)^T \vec{y} \right) \bullet \left(B^T A^T \vec{y} - (AB)^T \vec{y} \right) = 0$$

and so $B^T A^T \vec{y} - (AB)^T \vec{y} = \vec{0}$. However, this is true for all $\vec{y}$ and so $B^T A^T - (AB)^T = 0$.

4.7.8 $\frac{\left[\begin{array}{ccc} 3 & -1 & -1 \end{array}\right]^T \bullet \left[\begin{array}{ccc} 1 & 4 & 2 \end{array}\right]^T}{\sqrt{9+1+1}\sqrt{1+16+4}} = \frac{-3}{\sqrt{11}\sqrt{21}} = -0.197\,39 = \cos\theta$ Therefore we need to solve $-0.197\,39 = \cos\theta$. Thus $\theta = 1.7695$ radians.

4.7.9 $\frac{-10}{\sqrt{1+4+1}\sqrt{1+4+49}} = -0.55555 = \cos\theta$ Therefore we need to solve $-0.55555 = \cos\theta$, which gives $\theta = 2.031\,3$ radians.

4.7.10 $\frac{\vec{u}\bullet\vec{v}}{\vec{u}\bullet\vec{u}}\vec{u} = \frac{-5}{14}\left[\begin{array}{c} 1 \\ 2 \\ 3 \end{array}\right] = \left[\begin{array}{c} -\frac{5}{14} \\ -\frac{5}{7} \\ -\frac{15}{14} \end{array}\right]$

4.7.11 $\frac{\vec{u}\bullet\vec{v}}{\vec{u}\bullet\vec{u}}\vec{u} = \frac{-5}{10}\left[\begin{array}{c} 1 \\ 0 \\ 3 \end{array}\right] = \left[\begin{array}{c} -\frac{1}{2} \\ 0 \\ -\frac{3}{2} \end{array}\right]$

4.7.12 $\frac{\vec{u}\bullet\vec{v}}{\vec{u}\bullet\vec{u}}\vec{u} = \frac{\left[\begin{array}{cccc} 1 & 2 & -2 & 1 \end{array}\right]^T \bullet \left[\begin{array}{cccc} 1 & 2 & 3 & 0 \end{array}\right]^T}{1+4+9}\left[\begin{array}{c} 1 \\ 2 \\ 3 \\ 0 \end{array}\right] = \left[\begin{array}{c} -\frac{1}{14} \\ -\frac{1}{7} \\ -\frac{3}{14} \\ 0 \end{array}\right]$

4.7.15 No, it does not. The 0 vector has no direction. The formula for $\operatorname{proj}_{\vec{0}}(\vec{w})$ doesn't make sense either.

4.7.16

$$\left(\vec{u}-\frac{\vec{u}\bullet\vec{v}}{\|\vec{v}\|^2}\vec{v}\right)\bullet\left(\vec{u}-\frac{\vec{u}\bullet\vec{v}}{\|\vec{v}\|^2}\vec{v}\right)=\|\vec{u}\|^2-2\left(\vec{u}\bullet\vec{v}\right)^2\frac{1}{\|\vec{v}\|^2}+\left(\vec{u}\bullet\vec{v}\right)^2\frac{1}{\|\vec{v}\|^2}\geq 0$$

And so

$$\|\vec{u}\|^2\|\vec{v}\|^2\geq\left(\vec{u}\bullet\vec{v}\right)^2$$

You get equality exactly when $\vec{u}=\text{proj}_{\vec{v}}\vec{u}=\frac{\vec{u}\bullet\vec{v}}{\|\vec{v}\|^2}\vec{v}$ in other words, when $\vec{u}$ is a multiple of $\vec{v}$.

4.7.17

$$\begin{aligned}
&\vec{w}-\text{proj}_{\vec{v}}\left(\vec{w}\right)+\vec{u}-\text{proj}_{\vec{v}}\left(\vec{u}\right)\\
&\qquad=\vec{w}+\vec{u}-\left(\text{proj}_{\vec{v}}\left(\vec{w}\right)+\text{proj}_{\vec{v}}\left(\vec{u}\right)\right)\\
&\qquad=\vec{w}+\vec{u}-\text{proj}_{\vec{v}}\left(\vec{w}+\vec{u}\right)
\end{aligned}$$

This follows because

$$\begin{aligned}
\text{proj}_{\vec{v}}\left(\vec{w}\right)+\text{proj}_{\vec{v}}\left(\vec{u}\right) &= \frac{\vec{u}\bullet\vec{v}}{\|\vec{v}\|^2}\vec{v}+\frac{\vec{w}\bullet\vec{v}}{\|\vec{v}\|^2}\vec{v}\\
&= \frac{\left(\vec{u}+\vec{w}\right)\bullet\vec{v}}{\|\vec{v}\|^2}\vec{v}\\
&= \text{proj}_{\vec{v}}\left(\vec{w}+\vec{u}\right)
\end{aligned}$$

4.7.18 $\left(\vec{v}-\text{proj}_{\vec{u}}\left(\vec{v}\right)\right)\bullet\vec{u}=\vec{v}\bullet\vec{u}-\left(\frac{(v\cdot u)}{\|\vec{u}\|^2}\vec{u}\right)\bullet\vec{u}=\vec{v}\bullet\vec{u}-\vec{v}\bullet\vec{u}=0$. Therefore, $\vec{v}=\vec{v}-\text{proj}_{\vec{u}}\left(\vec{v}\right)+\text{proj}_{\vec{u}}\left(\vec{v}\right)$. The first is perpendicular to $\vec{u}$ and the second is a multiple of $\vec{u}$ so it is parallel to $\vec{u}$.

4.9.1 If $\vec{a}\neq\vec{0}$, then the condition says that $\|\vec{a}\times\vec{u}\|=\|\vec{a}\|\sin\theta=0$ for all angles θ. Hence $\vec{a}=\vec{0}$ after all.

4.9.2 $\left[\begin{array}{r}3\\0\\-3\end{array}\right]\times\left[\begin{array}{r}-4\\0\\-2\end{array}\right]=\left[\begin{array}{r}0\\18\\0\end{array}\right]$. So the area is 9.

4.9.3 $\left[\begin{array}{r}3\\1\\-3\end{array}\right]\times\left[\begin{array}{r}-4\\1\\-2\end{array}\right]=\left[\begin{array}{r}1\\18\\7\end{array}\right]$. The area is given by

$$\frac{1}{2}\sqrt{1+(18)^2+49}=\frac{1}{2}\sqrt{374}$$

4.9.4 $\left[\begin{array}{rrr}1&1&1\end{array}\right]\times\left[\begin{array}{rrr}2&2&2\end{array}\right]=\left[\begin{array}{rrr}0&0&0\end{array}\right]$. The area is 0. It means the three points are on the same line.

4.9.5 $\left[\begin{array}{r}1\\2\\3\end{array}\right]\times\left[\begin{array}{r}3\\-2\\1\end{array}\right]=\left[\begin{array}{r}8\\8\\-8\end{array}\right]$. The area is $8\sqrt{3}$

4.9.6 $\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} \times \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 11 \\ -2 \end{bmatrix}$. The area is $\sqrt{36+121+4} = \sqrt{161}$

4.9.7 $\left(\vec{i} \times \vec{j}\right) \times \vec{j} = \vec{k} \times \vec{j} = -\vec{i}$. However, $\vec{i} \times \left(\vec{j} \times \vec{j}\right) = \vec{0}$ and so the cross product is not associative.

4.9.8 Verify directly from the coordinate description of the cross product that the right hand rule applies to the vectors $\vec{i}, \vec{j}, \vec{k}$. Next verify that the distributive law holds for the coordinate description of the cross product. This gives another way to approach the cross product. First define it in terms of coordinates and then get the geometric properties from this. However, this approach does not yield the right hand rule property very easily. From the coordinate description,

$$\vec{a} \times \vec{b} \cdot \vec{a} = \varepsilon_{ijk} a_j b_k a_i = -\varepsilon_{jik} a_j b_k a_i = -\varepsilon_{jik} b_k a_i a_j = -\vec{a} \times \vec{b} \cdot \vec{a}$$

and so $\vec{a} \times \vec{b}$ is perpendicular to $\vec{a}$. Similarly, $\vec{a} \times \vec{b}$ is perpendicular to $\vec{b}$. Now we need that

$$\|\vec{a} \times \vec{b}\|^2 = \|\vec{a}\|^2 \|\vec{b}\|^2 \left(1 - \cos^2 \theta\right) = \|\vec{a}\|^2 \|\vec{b}\|^2 \sin^2 \theta$$

and so $\|\vec{a} \times \vec{b}\| = \|\vec{a}\| \|\vec{b}\| \sin \theta$, the area of the parallelogram determined by $\vec{a}, \vec{b}$. Only the right hand rule is a little problematic. However, you can see right away from the component definition that the right hand rule holds for each of the standard unit vectors. Thus $\vec{i} \times \vec{j} = \vec{k}$ etc.

$$\begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = \vec{k}$$

4.9.10 $\begin{vmatrix} 1 & -7 & -5 \\ 1 & -2 & -6 \\ 3 & 2 & 3 \end{vmatrix} = 113$

4.9.11 Yes. It will involve the sum of product of integers and so it will be an integer.

4.9.12 It means that if you place them so that they all have their tails at the same point, the three will lie in the same plane.

4.9.13 $\vec{x} \bullet \left(\vec{a} \times \vec{b}\right) = 0$

4.9.15 Here $[\vec{v}, \vec{w}, \vec{z}]$ denotes the box product. Consider the cross product term. From the above,

$$\begin{aligned} (\vec{v} \times \vec{w}) \times (\vec{w} \times \vec{z}) &= [\vec{v}, \vec{w}, \vec{z}]\, \vec{w} - [\vec{w}, \vec{w}, \vec{z}]\, \vec{v} \\ &= [\vec{v}, \vec{w}, \vec{z}]\, \vec{w} \end{aligned}$$

Thus it reduces to

$$(\vec{u} \times \vec{v}) \bullet [\vec{v}, \vec{w}, \vec{z}]\, \vec{w} = [\vec{v}, \vec{w}, \vec{z}]\, [\vec{u}, \vec{v}, \vec{w}]$$

4.9.16

$$\begin{aligned} \|\vec{u} \times \vec{v}\|^2 &= \varepsilon_{ijk} u_j v_k \varepsilon_{irs} u_r v_s = (\delta_{jr}\delta_{ks} - \delta_{kr}\delta_{js})\, u_r v_s u_j v_k \\ &= u_j v_k u_j v_k - u_k v_j u_j v_k = \|\vec{u}\|^2 \|\vec{v}\|^2 - (\vec{u} \bullet \vec{v})^2 \end{aligned}$$

It follows that the expression reduces to 0. You can also do the following.

$$\begin{aligned} \|\vec{u} \times \vec{v}\|^2 &= \|\vec{u}\|^2 \|\vec{v}\|^2 \sin^2 \theta \\ &= \|\vec{u}\|^2 \|\vec{v}\|^2 \left(1 - \cos^2 \theta\right) \\ &= \|\vec{u}\|^2 \|\vec{v}\|^2 - \|\vec{u}\|^2 \|\vec{v}\|^2 \cos^2 \theta \\ &= \|\vec{u}\|^2 \|\vec{v}\|^2 - (\vec{u} \bullet \vec{v})^2 \end{aligned}$$

which implies the expression equals 0.

4.9.17 We will show it using the summation convention and permutation symbol.

$$\begin{aligned} \left((\vec{u} \times \vec{v})'\right)_i &= \left((\vec{u} \times \vec{v})_i\right)' = (\varepsilon_{ijk} u_j v_k)' \\ &= \varepsilon_{ijk} u_j' v_k + \varepsilon_{ijk} u_k v_k' = (\vec{u}' \times \vec{v} + \vec{u} \times \vec{v}')_i \end{aligned}$$

and so $(\vec{u} \times \vec{v})' = \vec{u}' \times \vec{v} + \vec{u} \times \vec{v}'$.

4.10.10 $\sum_{i=1}^{k} 0\vec{x}_k = \vec{0}$

4.10.40 No. Let $\vec{u} = \left[\begin{array}{c} \frac{\pi}{2} \\ 0 \\ 0 \\ 0 \end{array}\right]$. Then $2\vec{u} \notin M$ although $\vec{u} \in M$.

4.10.41 No. $\left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \end{array}\right] \in M$ but $10 \left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \end{array}\right] \notin M$.

4.10.42 This is not a subspace. $\left[\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \end{array}\right]$ is in it. However, $(-1) \left[\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \end{array}\right]$ is not.

4.10.43 This is a subspace because it is closed with respect to vector addition and scalar multiplication.

4.10.44 Yes, this is a subspace because it is closed with respect to vector addition and scalar multiplication.

4.10.45 This is not a subspace. $\left[\begin{array}{c} 0 \\ 0 \\ 1 \\ 0 \end{array}\right]$ is in it. However $(-1) \left[\begin{array}{c} 0 \\ 0 \\ 1 \\ 0 \end{array}\right] = \left[\begin{array}{r} 0 \\ 0 \\ -1 \\ 0 \end{array}\right]$ is not.

4.10.46 This is a subspace. It is closed with respect to vector addition and scalar multiplication.

4.10.55 Yes. If not, there would exist a vector not in the span. But then you could add in this vector and obtain a linearly independent set of vectors with more vectors than a basis.

4.10.56 They can't be.

4.10.57 Say $\sum_{i=1}^{k} c_i \vec{z}_i = \vec{0}$. Then apply A to it as follows.

$$\sum_{i=1}^{k} c_i A \vec{z}_i = \sum_{i=1}^{k} c_i \vec{w}_i = \vec{0}$$

and so, by linear independence of the $\vec{w}_i$, it follows that each $c_i = 0$.

4.10.58 If $\vec{x}, \vec{y} \in V \cap W$, then for scalars α, β, the linear combination $\alpha\vec{x} + \beta\vec{y}$ must be in both V and W since they are both subspaces.

4.10.60 Let $\{x_1, \cdots, x_k\}$ be a basis for $V \cap W$. Then there is a basis for V and W which are respectively

$$\{x_1, \cdots, x_k, y_{k+1}, \cdots, y_p\}, \ \{x_1, \cdots, x_k, z_{k+1}, \cdots, z_q\}$$

It follows that you must have $k + p - k + q - k \leq n$ and so you must have

$$p + q - n \leq k$$

4.10.61 Here is how you do this. Suppose $AB\vec{x} = \vec{0}$. Then $B\vec{x} \in \ker(A) \cap B(\mathbb{R}^p)$ and so $B\vec{x} = \sum_{i=1}^{k} B\vec{z}_i$ showing that

$$\vec{x} - \sum_{i=1}^{k} \vec{z}_i \in \ker(B)$$

Consider $B(\mathbb{R}^p) \cap \ker(A)$ and let a basis be $\{\vec{w}_1, \cdots, \vec{w}_k\}$. Then each $\vec{w}_i$ is of the form $B\vec{z}_i = \vec{w}_i$. Therefore, $\{\vec{z}_1, \cdots, \vec{z}_k\}$ is linearly independent and $AB\vec{z}_i = 0$. Now let $\{\vec{u}_1, \cdots, \vec{u}_r\}$ be a basis for $\ker(B)$. If $AB\vec{x} = \vec{0}$, then $B\vec{x} \in \ker(A) \cap B(\mathbb{R}^p)$ and so $B\vec{x} = \sum_{i=1}^{k} c_i B\vec{z}_i$ which implies

$$\vec{x} - \sum_{i=1}^{k} c_i \vec{z}_i \in \ker(B)$$

and so it is of the form

$$\vec{x} - \sum_{i=1}^{k} c_i \vec{z}_i = \sum_{j=1}^{r} d_j \vec{u}_j$$

It follows that if $AB\vec{x} = \vec{0}$ so that $\vec{x} \in \ker(AB)$, then

$$\vec{x} \in \operatorname{span}(\vec{z}_1, \cdots, \vec{z}_k, \vec{u}_1, \cdots, \vec{u}_r).$$

Therefore,

$$\begin{aligned} \dim(\ker(AB)) &\leq k + r = \dim(B(\mathbb{R}^p) \cap \ker(A)) + \dim(\ker(B)) \\ &\leq \dim(\ker(A)) + \dim(\ker(B)) \end{aligned}$$

4.10.62 If $\vec{x}, \vec{y} \in \ker(A)$ then

$$A(a\vec{x} + b\vec{y}) = aA\vec{x} + bA\vec{y} = a\vec{0} + b\vec{0} = \vec{0}$$

and so $\ker(A)$ is closed under linear combinations. Hence it is a subspace.

4.11.6
1. Orthogonal.
2. Symmetric.
3. Skew Symmetric.

4.11.7 $\|U\vec{x}\|^2 = U\vec{x} \bullet U\vec{x} = U^T U\vec{x} \bullet \vec{x} = I\vec{x} \bullet \vec{x} = \|\vec{x}\|^2$.

Next suppose distance is preserved by U. Then

$$\begin{aligned}(U(\vec{x}+\vec{y})) \bullet (U(\vec{x}+\vec{y})) &= \|Ux\|^2 + \|Uy\|^2 + 2(Ux \bullet Uy) \\ &= \|\vec{x}\|^2 + \|\vec{y}\|^2 + 2\left(U^T U\vec{x} \bullet \vec{y}\right)\end{aligned}$$

But since U preserves distances, it is also the case that

$$(U(\vec{x}+\vec{y}) \bullet U(\vec{x}+\vec{y})) = \|\vec{x}\|^2 + \|\vec{y}\|^2 + 2(\vec{x} \bullet \vec{y})$$

Hence

$$\vec{x} \bullet \vec{y} = U^T U\vec{x} \bullet \vec{y}$$

and so

$$\left(\left(U^T U - I\right)\vec{x}\right) \bullet \vec{y} = 0$$

Since y is arbitrary, it follows that $U^T U - I = 0$. Thus U is orthogonal.

4.11.8 You could observe that $\det\left(UU^T\right) = (\det(U))^2 = 1$ so $\det(U) \neq 0$.

4.11.9

$$\left[\begin{array}{ccc} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & a \\ 0 & \frac{\sqrt{6}}{3} & b \end{array}\right]\left[\begin{array}{ccc} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & a \\ 0 & \frac{\sqrt{6}}{3} & b \end{array}\right]^T$$

$$= \left[\begin{array}{ccc} 1 & \frac{1}{3}\sqrt{3}a - \frac{1}{3} & \frac{1}{3}\sqrt{3}b - \frac{1}{3} \\ \frac{1}{3}\sqrt{3}a - \frac{1}{3} & a^2 + \frac{2}{3} & ab - \frac{1}{3} \\ \frac{1}{3}\sqrt{3}b - \frac{1}{3} & ab - \frac{1}{3} & b^2 + \frac{2}{3} \end{array}\right]$$

This requires $a = 1/\sqrt{3}, b = 1/\sqrt{3}$.

$$\left[\begin{array}{ccc} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & 1/\sqrt{3} \\ 0 & \frac{\sqrt{6}}{3} & 1/\sqrt{3} \end{array}\right]\left[\begin{array}{ccc} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & 1/\sqrt{3} \\ 0 & \frac{\sqrt{6}}{3} & 1/\sqrt{3} \end{array}\right]^T = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$$

4.11.10 $\begin{bmatrix} \frac{2}{3} & \frac{\sqrt{2}}{2} & \frac{1}{6}\sqrt{2} \\ \frac{2}{3} & \frac{-\sqrt{2}}{2} & a \\ -\frac{1}{3} & 0 & b \end{bmatrix} \begin{bmatrix} \frac{2}{3} & \frac{\sqrt{2}}{2} & \frac{1}{6}\sqrt{2} \\ \frac{2}{3} & \frac{-\sqrt{2}}{2} & a \\ -\frac{1}{3} & 0 & b \end{bmatrix}^T = \begin{bmatrix} 1 & \frac{1}{6}\sqrt{2}a - \frac{1}{18} & \frac{1}{6}\sqrt{2}b - \frac{2}{9} \\ \frac{1}{6}\sqrt{2}a - \frac{1}{18} & a^2 + \frac{17}{18} & ab - \frac{2}{9} \\ \frac{1}{6}\sqrt{2}b - \frac{2}{9} & ab - \frac{2}{9} & b^2 + \frac{1}{9} \end{bmatrix}$

This requires $a = \frac{1}{3\sqrt{2}}, b = \frac{4}{3\sqrt{2}}$.

$$\begin{bmatrix} \frac{2}{3} & \frac{\sqrt{2}}{2} & \frac{1}{6}\sqrt{2} \\ \frac{2}{3} & \frac{-\sqrt{2}}{2} & \frac{1}{3\sqrt{2}} \\ -\frac{1}{3} & 0 & \frac{4}{3\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{2}{3} & \frac{\sqrt{2}}{2} & \frac{1}{6}\sqrt{2} \\ \frac{2}{3} & \frac{-\sqrt{2}}{2} & \frac{1}{3\sqrt{2}} \\ -\frac{1}{3} & 0 & \frac{4}{3\sqrt{2}} \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

4.11.11 Try

$$\begin{bmatrix} \frac{1}{3} & -\frac{2}{\sqrt{5}} & c \\ \frac{2}{3} & 0 & d \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{15}\sqrt{5} \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{2}{\sqrt{5}} & c \\ \frac{2}{3} & 0 & d \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{15}\sqrt{5} \end{bmatrix}^T$$

$$= \begin{bmatrix} c^2 + \frac{41}{45} & cd + \frac{2}{9} & \frac{4}{15}\sqrt{5}c - \frac{8}{45} \\ cd + \frac{2}{9} & d^2 + \frac{4}{9} & \frac{4}{15}\sqrt{5}d + \frac{4}{9} \\ \frac{4}{15}\sqrt{5}c - \frac{8}{45} & \frac{4}{15}\sqrt{5}d + \frac{4}{9} & 1 \end{bmatrix}$$

This requires that $c = \frac{2}{3\sqrt{5}}, d = \frac{-5}{3\sqrt{5}}$.

$$\begin{bmatrix} \frac{1}{3} & -\frac{2}{\sqrt{5}} & \frac{2}{3\sqrt{5}} \\ \frac{2}{3} & 0 & \frac{-5}{3\sqrt{5}} \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{15}\sqrt{5} \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{2}{\sqrt{5}} & \frac{2}{3\sqrt{5}} \\ \frac{2}{3} & 0 & \frac{-5}{3\sqrt{5}} \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{15}\sqrt{5} \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

4.11.12

1. $$\begin{bmatrix} \frac{3}{5} \\ -\frac{4}{5} \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{4}{5} \\ \frac{3}{5} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

2. $$\begin{bmatrix} \frac{3}{5} \\ 0 \\ -\frac{4}{5} \end{bmatrix}, \begin{bmatrix} \frac{4}{5} \\ 0 \\ \frac{3}{5} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

3. $$\begin{bmatrix} \frac{3}{5} \\ 0 \\ -\frac{4}{5} \end{bmatrix}, \begin{bmatrix} \frac{4}{5} \\ 0 \\ \frac{3}{5} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

4.11.13 A solution is

$$\begin{bmatrix} \frac{1}{6}\sqrt{6} \\ \frac{1}{3}\sqrt{6} \\ \frac{1}{6}\sqrt{6} \end{bmatrix}, \begin{bmatrix} \frac{3}{10}\sqrt{2} \\ -\frac{2}{5}\sqrt{2} \\ \frac{1}{2}\sqrt{2} \end{bmatrix}, \begin{bmatrix} \frac{7}{15}\sqrt{3} \\ -\frac{1}{15}\sqrt{3} \\ -\frac{1}{3}\sqrt{3} \end{bmatrix}$$

4.11.14 Then a solution is

$$\begin{bmatrix} \frac{1}{6}\sqrt{6} \\ \frac{1}{3}\sqrt{6} \\ \frac{1}{6}\sqrt{6} \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{6}\sqrt{2}\sqrt{3} \\ -\frac{2}{9}\sqrt{2}\sqrt{3} \\ \frac{5}{18}\sqrt{2}\sqrt{3} \\ \frac{1}{9}\sqrt{2}\sqrt{3} \end{bmatrix}, \begin{bmatrix} \frac{5}{111}\sqrt{3}\sqrt{37} \\ \frac{1}{333}\sqrt{3}\sqrt{37} \\ -\frac{17}{333}\sqrt{3}\sqrt{37} \\ \frac{22}{333}\sqrt{3}\sqrt{37} \end{bmatrix}$$

4.11.15 The subspace is of the form

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

and a basis is $\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$. Therefore, an orthonormal basis is

$$\begin{bmatrix} \frac{1}{5}\sqrt{5} \\ 0 \\ \frac{2}{5}\sqrt{5} \end{bmatrix}, \begin{bmatrix} -\frac{3}{35}\sqrt{5}\sqrt{14} \\ \frac{1}{14}\sqrt{5}\sqrt{14} \\ \frac{3}{70}\sqrt{5}\sqrt{14} \end{bmatrix}$$

5.1.1 This result follows from the properties of matrix multiplication.

5.1.2

$$\begin{aligned} T_{\vec{u}}\left(a\vec{v}+b\vec{w}\right) &= a\vec{v}+b\vec{w}-\frac{(a\vec{v}+b\vec{w}\bullet\vec{u})}{\|\vec{u}\|^2}\vec{u} \\ &= a\vec{v}-a\frac{(\vec{v}\bullet\vec{u})}{\|\vec{u}\|^2}\vec{u}+b\vec{w}-b\frac{(\vec{w}\bullet\vec{u})}{\|\vec{u}\|^2}\vec{u} \\ &= aT_{\vec{u}}\left(\vec{v}\right)+bT_{\vec{u}}\left(\vec{w}\right) \end{aligned}$$

5.1.3 Linear transformations take $\vec{0}$ to $\vec{0}$ which T does not. Also $T_{\vec{a}}\left(\vec{u}+\vec{v}\right) \neq T_{\vec{a}}\vec{u}+T_{\vec{a}}\vec{v}$.

5.2.1

1. The matrix of T is the elementary matrix which multiplies the j^{th} diagonal entry of the identity matrix by b.
2. The matrix of T is the elementary matrix which takes b times the j^{th} row and adds to the i^{th} row.
3. The matrix of T is the elementary matrix which switches the i^{th} and the j^{th} rows where the two components are in the i^{th} and j^{th} positions.

5.2.2 Suppose

$$\begin{bmatrix} \vec{c}_1^T \\ \vdots \\ \vec{c}_n^T \end{bmatrix} = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix}^{-1}$$

Thus $\vec{c}_i^T \vec{a}_j = \delta_{ij}$. Therefore,

$$\begin{aligned}
\left[\begin{array}{ccc} \vec{b}_1 & \cdots & \vec{b}_n \end{array} \right] \left[\begin{array}{ccc} \vec{a}_1 & \cdots & \vec{a}_n \end{array} \right]^{-1} \vec{a}_i &= \left[\begin{array}{ccc} \vec{b}_1 & \cdots & \vec{b}_n \end{array} \right] \left[\begin{array}{c} \vec{c}_1^T \\ \vdots \\ \vec{c}_n^T \end{array} \right] \vec{a}_i \\
&= \left[\begin{array}{ccc} \vec{b}_1 & \cdots & \vec{b}_n \end{array} \right] \vec{e}_i \\
&= \vec{b}_i
\end{aligned}$$

Thus $T\vec{a}_i = \left[\begin{array}{ccc} \vec{b}_1 & \cdots & \vec{b}_n \end{array} \right] \left[\begin{array}{ccc} \vec{a}_1 & \cdots & \vec{a}_n \end{array} \right]^{-1} \vec{a}_i = A\vec{a}_i$. If $\vec{x}$ is arbitrary, then since the matrix $\left[\begin{array}{ccc} \vec{a}_1 & \cdots & \vec{a}_n \end{array} \right]$ is invertible, there exists a unique $\vec{y}$ such that $\left[\begin{array}{ccc} \vec{a}_1 & \cdots & \vec{a}_n \end{array} \right] \vec{y} = \vec{x}$ Hence

$$T\vec{x} = T\left(\sum_{i=1}^{n} y_i \vec{a}_i \right) = \sum_{i=1}^{n} y_i T\vec{a}_i = \sum_{i=1}^{n} y_i A\vec{a}_i = A\left(\sum_{i=1}^{n} y_i \vec{a}_i \right) = A\vec{x}$$

5.2.3

$$\left[\begin{array}{rrr} 5 & 1 & 5 \\ 1 & 1 & 3 \\ 3 & 5 & -2 \end{array} \right] \left[\begin{array}{rrr} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 4 & 1 & 1 \end{array} \right] = \left[\begin{array}{rrr} 37 & 17 & 11 \\ 17 & 7 & 5 \\ 11 & 14 & 6 \end{array} \right]$$

5.2.4

$$\left[\begin{array}{rrr} 1 & 2 & 6 \\ 3 & 4 & 1 \\ 1 & 1 & -1 \end{array} \right] \left[\begin{array}{rrr} 6 & 3 & 1 \\ 5 & 3 & 1 \\ 6 & 2 & 1 \end{array} \right] = \left[\begin{array}{rrr} 52 & 21 & 9 \\ 44 & 23 & 8 \\ 5 & 4 & 1 \end{array} \right]$$

5.2.5

$$\left[\begin{array}{rrr} -3 & 1 & 5 \\ 1 & 3 & 3 \\ 3 & -3 & -3 \end{array} \right] \left[\begin{array}{rrr} 2 & 2 & 1 \\ 1 & 2 & 1 \\ 4 & 1 & 1 \end{array} \right] = \left[\begin{array}{rrr} 15 & 1 & 3 \\ 17 & 11 & 7 \\ -9 & -3 & -3 \end{array} \right]$$

5.2.6

$$\left[\begin{array}{rrr} 3 & 1 & 1 \\ 3 & 2 & 3 \\ 3 & 3 & -1 \end{array} \right] \left[\begin{array}{rrr} 6 & 2 & 1 \\ 5 & 2 & 1 \\ 6 & 1 & 1 \end{array} \right] = \left[\begin{array}{rrr} 29 & 9 & 5 \\ 46 & 13 & 8 \\ 27 & 11 & 5 \end{array} \right]$$

5.2.7

$$\left[\begin{array}{rrr} 5 & 3 & 2 \\ 2 & 3 & 5 \\ 5 & 5 & -2 \end{array} \right] \left[\begin{array}{rrr} 11 & 4 & 1 \\ 10 & 4 & 1 \\ 12 & 3 & 1 \end{array} \right] = \left[\begin{array}{rrr} 109 & 38 & 10 \\ 112 & 35 & 10 \\ 81 & 34 & 8 \end{array} \right]$$

5.2.11 Recall that $\text{proj}_{\vec{u}}(\vec{v}) = \frac{\vec{v}\bullet\vec{u}}{\|\vec{u}\|^2}\vec{u}$ and so the desired matrix has i^{th} column equal to $\text{proj}_{\vec{u}}(\vec{e}_i)$. Therefore, the matrix desired is

$$\frac{1}{14} \left[\begin{array}{rrr} 1 & -2 & 3 \\ -2 & 4 & -6 \\ 3 & -6 & 9 \end{array} \right]$$

5.2.12

$$\frac{1}{35}\begin{bmatrix} 1 & 5 & 3 \\ 5 & 25 & 15 \\ 3 & 15 & 9 \end{bmatrix}$$

5.2.13

$$\frac{1}{10}\begin{bmatrix} 1 & 0 & 3 \\ 0 & 0 & 0 \\ 3 & 0 & 9 \end{bmatrix}$$

5.3.2 The matrix of $S \circ T$ is given by BA.

$$\begin{bmatrix} 0 & -2 \\ 4 & 2 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & -4 \\ 10 & 8 \end{bmatrix}$$

Now, $(S \circ T)(\vec{x}) = (BA)\vec{x}$.

$$\begin{bmatrix} 2 & -4 \\ 10 & 8 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 8 \\ 12 \end{bmatrix}$$

5.3.3 To find $(S \circ T)(\vec{x})$ we compute $S(T(\vec{x}))$.

$$\begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 2 \\ -3 \end{bmatrix} = \begin{bmatrix} -4 \\ -11 \end{bmatrix}$$

5.3.5 The matrix of T^{-1} is A^{-1}.

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

5.4.1 $\begin{bmatrix} \cos\left(\frac{\pi}{3}\right) & -\sin\left(\frac{\pi}{3}\right) \\ \sin\left(\frac{\pi}{3}\right) & \cos\left(\frac{\pi}{3}\right) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & \frac{1}{2} \end{bmatrix}$

5.4.2 $\begin{bmatrix} \cos\left(\frac{\pi}{4}\right) & -\sin\left(\frac{\pi}{4}\right) \\ \sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{bmatrix}$

5.4.3 $\begin{bmatrix} \cos\left(-\frac{\pi}{3}\right) & -\sin\left(-\frac{\pi}{3}\right) \\ \sin\left(-\frac{\pi}{3}\right) & \cos\left(-\frac{\pi}{3}\right) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} & \frac{1}{2} \end{bmatrix}$

5.4.4 $\begin{bmatrix} \cos\left(\frac{2\pi}{3}\right) & -\sin\left(\frac{2\pi}{3}\right) \\ \sin\left(\frac{2\pi}{3}\right) & \cos\left(\frac{2\pi}{3}\right) \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{bmatrix}$

5.4.5

$$\begin{bmatrix} \cos\left(\frac{\pi}{3}\right) & -\sin\left(\frac{\pi}{3}\right) \\ \sin\left(\frac{\pi}{3}\right) & \cos\left(\frac{\pi}{3}\right) \end{bmatrix}\begin{bmatrix} \cos\left(-\frac{\pi}{4}\right) & -\sin\left(-\frac{\pi}{4}\right) \\ \sin\left(-\frac{\pi}{4}\right) & \cos\left(-\frac{\pi}{4}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{4}\sqrt{2}\sqrt{3}+\frac{1}{4}\sqrt{2} & \frac{1}{4}\sqrt{2}-\frac{1}{4}\sqrt{2}\sqrt{3} \\ \frac{1}{4}\sqrt{2}\sqrt{3}-\frac{1}{4}\sqrt{2} & \frac{1}{4}\sqrt{2}\sqrt{3}+\frac{1}{4}\sqrt{2} \end{bmatrix}$$

5.4.6

$$\left[\begin{array}{rr} 1 & 0 \\ 0 & -1 \end{array}\right]\left[\begin{array}{cc} \cos\left(\frac{2\pi}{3}\right) & -\sin\left(\frac{2\pi}{3}\right) \\ \sin\left(\frac{2\pi}{3}\right) & \cos\left(\frac{2\pi}{3}\right) \end{array}\right] = \left[\begin{array}{cc} -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} & \frac{1}{2} \end{array}\right]$$

5.4.7

$$\left[\begin{array}{rr} 1 & 0 \\ 0 & -1 \end{array}\right]\left[\begin{array}{cc} \cos\left(\frac{\pi}{3}\right) & -\sin\left(\frac{\pi}{3}\right) \\ \sin\left(\frac{\pi}{3}\right) & \cos\left(\frac{\pi}{3}\right) \end{array}\right] = \left[\begin{array}{cc} \frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{array}\right]$$

5.4.8

$$\left[\begin{array}{rr} 1 & 0 \\ 0 & -1 \end{array}\right]\left[\begin{array}{cc} \cos\left(\frac{\pi}{4}\right) & -\sin\left(\frac{\pi}{4}\right) \\ \sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{array}\right] = \left[\begin{array}{cc} \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \end{array}\right]$$

5.4.9

$$\left[\begin{array}{rr} -1 & 0 \\ 0 & 1 \end{array}\right]\left[\begin{array}{cc} \cos\left(\frac{\pi}{6}\right) & -\sin\left(\frac{\pi}{6}\right) \\ \sin\left(\frac{\pi}{6}\right) & \cos\left(\frac{\pi}{6}\right) \end{array}\right] = \left[\begin{array}{cc} -\frac{1}{2}\sqrt{3} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}\sqrt{3} \end{array}\right]$$

5.4.10

$$\left[\begin{array}{cc} \cos\left(\frac{\pi}{4}\right) & -\sin\left(\frac{\pi}{4}\right) \\ \sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{array}\right]\left[\begin{array}{rr} 1 & 0 \\ 0 & -1 \end{array}\right] = \left[\begin{array}{cc} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \end{array}\right]$$

5.4.11

$$\left[\begin{array}{cc} \cos\left(\frac{\pi}{4}\right) & -\sin\left(\frac{\pi}{4}\right) \\ \sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{array}\right]\left[\begin{array}{rr} -1 & 0 \\ 0 & 1 \end{array}\right] = \left[\begin{array}{cc} -\frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{array}\right]$$

5.4.12

$$\left[\begin{array}{cc} \cos\left(\frac{\pi}{6}\right) & -\sin\left(\frac{\pi}{6}\right) \\ \sin\left(\frac{\pi}{6}\right) & \cos\left(\frac{\pi}{6}\right) \end{array}\right]\left[\begin{array}{rr} 1 & 0 \\ 0 & -1 \end{array}\right] = \left[\begin{array}{cc} \frac{1}{2}\sqrt{3} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2}\sqrt{3} \end{array}\right]$$

5.4.13

$$\left[\begin{array}{cc} \cos\left(\frac{\pi}{6}\right) & -\sin\left(\frac{\pi}{6}\right) \\ \sin\left(\frac{\pi}{6}\right) & \cos\left(\frac{\pi}{6}\right) \end{array}\right]\left[\begin{array}{rr} -1 & 0 \\ 0 & 1 \end{array}\right] = \left[\begin{array}{cc} -\frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2}\sqrt{3} \end{array}\right]$$

5.4.14

$$\left[\begin{array}{cc} \cos\left(\frac{2\pi}{3}\right) & -\sin\left(\frac{2\pi}{3}\right) \\ \sin\left(\frac{2\pi}{3}\right) & \cos\left(\frac{2\pi}{3}\right) \end{array}\right]\left[\begin{array}{cc} \cos\left(-\frac{\pi}{4}\right) & -\sin\left(-\frac{\pi}{4}\right) \\ \sin\left(-\frac{\pi}{4}\right) & \cos\left(-\frac{\pi}{4}\right) \end{array}\right] =$$

$$\left[\begin{array}{cc} \frac{1}{4}\sqrt{2}\sqrt{3}-\frac{1}{4}\sqrt{2} & -\frac{1}{4}\sqrt{2}\sqrt{3}-\frac{1}{4}\sqrt{2} \\ \frac{1}{4}\sqrt{2}\sqrt{3}+\frac{1}{4}\sqrt{2} & \frac{1}{4}\sqrt{2}\sqrt{3}-\frac{1}{4}\sqrt{2} \end{array}\right]$$

Note that it doesn't matter about the order in this case.

5.4.15

$$\left[\begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{array}\right]\left[\begin{array}{ccc} \cos\left(\frac{\pi}{6}\right) & -\sin\left(\frac{\pi}{6}\right) & 0 \\ \sin\left(\frac{\pi}{6}\right) & \cos\left(\frac{\pi}{6}\right) & 0 \\ 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{ccc} \frac{1}{2}\sqrt{3} & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2}\sqrt{3} & 0 \\ 0 & 0 & -1 \end{array}\right]$$

5.4.16

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix}$$
$$= \begin{bmatrix} \cos^2\theta - \sin^2\theta & 2\cos\theta\sin\theta \\ 2\cos\theta\sin\theta & \sin^2\theta - \cos^2\theta \end{bmatrix}$$

Now to write in terms of (a, b), note that $a/\sqrt{a^2+b^2} = \cos\theta, b/\sqrt{a^2+b^2} = \sin\theta$. Now plug this in to the above. The result is

$$\begin{bmatrix} \frac{a^2-b^2}{a^2+b^2} & 2\frac{ab}{a^2+b^2} \\ 2\frac{ab}{a^2+b^2} & \frac{b^2-a^2}{a^2+b^2} \end{bmatrix} = \frac{1}{a^2+b^2} \begin{bmatrix} a^2-b^2 & 2ab \\ 2ab & b^2-a^2 \end{bmatrix}$$

Since this is a unit vector, $a^2 + b^2 = 1$ and so you get

$$\begin{bmatrix} a^2-b^2 & 2ab \\ 2ab & b^2-a^2 \end{bmatrix}$$

5.5.5 $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$

5.5.6 This says that the columns of A have a subset of m vectors which are linearly independent. Therefore, this set of vectors is a basis for $\mathbb{R}^m$. It follows that the span of the columns is all of $\mathbb{R}^m$. Thus A is onto.

5.5.7 The columns are independent. Therefore, A is one to one.

5.5.8 The rank is n is the same as saying the columns are independent which is the same as saying A is one to one which is the same as saying the columns are a basis. Thus the span of the columns of A is all of $\mathbb{R}^n$ and so A is onto. If A is onto, then the columns must be linearly independent since otherwise the span of these columns would have dimension less than n and so the dimension of $\mathbb{R}^n$ would be less than n .

5.6.1 Solution is: $\begin{bmatrix} -3\hat{t} \\ -\hat{t} \\ \hat{t} \end{bmatrix}, \hat{t}_3 \in \mathbb{R}$. A basis for the solution space is $\begin{bmatrix} -3 \\ -1 \\ 1 \end{bmatrix}$

5.6.2 Note that this has the same matrix as the above problem. Solution is: $\begin{bmatrix} -3\hat{t}_3 \\ -\hat{t}_3 \\ \hat{t}_3 \end{bmatrix} +$ $\begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \hat{t}_3 \in \mathbb{R}$

5.6.3 Solution is: $\begin{bmatrix} 3\hat{t} \\ 2\hat{t} \\ \hat{t} \end{bmatrix}$, A basis is $\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$

5.6.4 Solution is: $\begin{bmatrix} 3\hat{t} \\ 2\hat{t} \\ \hat{t} \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ 0 \end{bmatrix}$, $\hat{t} \in \mathbb{R}$

5.6.5 Solution is: $\begin{bmatrix} -4\hat{t} \\ -2\hat{t} \\ \hat{t} \end{bmatrix}$. A basis is $\begin{bmatrix} -4 \\ -2 \\ 1 \end{bmatrix}$

5.6.6 Solution is: $\begin{bmatrix} -4\hat{t} \\ -2\hat{t} \\ \hat{t} \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$, $\hat{t} \in \mathbb{R}$.

5.6.7 Solution is: $\begin{bmatrix} -\hat{t} \\ 2\hat{t} \\ \hat{t} \end{bmatrix}$, $\hat{t} \in \mathbb{R}$.

5.6.8 Solution is: $\begin{bmatrix} -\hat{t} \\ 2\hat{t} \\ \hat{t} \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$

5.6.9 Solution is: $\begin{bmatrix} 0 \\ -\hat{t} \\ -\hat{t} \\ \hat{t} \end{bmatrix}$, $\hat{t} \in \mathbb{R}$

5.6.10 Solution is: $\begin{bmatrix} 0 \\ -\hat{t} \\ -\hat{t} \\ \hat{t} \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \\ -1 \\ 0 \end{bmatrix}$

5.6.11 Solution is: $\begin{bmatrix} -s-t \\ s \\ s \\ t \end{bmatrix}$, $s, t \in \mathbb{R}$. A basis is

$$\left\{ \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

5.6.12 Solution is:

$$\begin{bmatrix} -\hat{t} \\ \hat{t} \\ \hat{t} \\ 0 \end{bmatrix} + \begin{bmatrix} -8 \\ 5 \\ 0 \\ 5 \end{bmatrix}$$

5.6.13 Solution is:
$$\begin{bmatrix} -\frac{1}{2}s - \frac{1}{2}t \\ \frac{1}{2}s - \frac{1}{2}t \\ s \\ t \end{bmatrix}$$
for $s, t \in \mathbb{R}$. A basis is
$$\left\{ \begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

5.6.14 Solution is:
$$\begin{bmatrix} \frac{3}{2} \\ -\frac{1}{2} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -\frac{1}{2}s - \frac{1}{2}t \\ \frac{1}{2}s - \frac{1}{2}t \\ s \\ t \end{bmatrix}$$

5.6.15 Solution is: $\begin{bmatrix} -\hat{t} \\ \hat{t} \\ \hat{t} \\ 0 \end{bmatrix}$, a basis is $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$.

5.6.16 Solution is: $\begin{bmatrix} -\hat{t} \\ \hat{t} \\ \hat{t} \\ 0 \end{bmatrix} + \begin{bmatrix} -9 \\ 5 \\ 0 \\ 6 \end{bmatrix}, t \in \mathbb{R}$.

5.6.17 If not, then there would be a infintely many solutions to $A\vec{x} = \vec{0}$ and each of these added to a solution to $A\vec{x} = \vec{b}$ would be a solution to $A\vec{x} = \vec{b}$.

6.1.1 $A^m X = \lambda^m X$ for any integer. In the case of $-1, A^{-1}\lambda X = AA^{-1}X = X$ so $A^{-1}X = \lambda^{-1}X$. Thus the eigenvalues of A^{-1} are just λ^{-1} where λ is an eigenvalue of A.

6.1.2 Say $AX = \lambda X$. Then $cAX = c\lambda X$ and so the eigenvalues of cA are just $c\lambda$ where λ is an eigenvalue of A.

6.1.3 $BAX = ABX = A\lambda X = \lambda AX$. Here it is assumed that $BX = \lambda X$.

6.1.4 Let X be the eigenvector. Then $A^m X = \lambda^m X, A^m X = AX = \lambda X$ and so
$$\lambda^m = \lambda$$
Hence if $\lambda \neq 0$, then
$$\lambda^{m-1} = 1$$
and so $|\lambda| = 1$.

6.1.5 The formula follows from properties of matrix multiplications. However, this vector might not be an eigenvector because it might equal 0 and eigenvectors cannot equal 0.

6.1.14 Yes. $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ works.

6.2.1 The eigenvalues are $-1, -1, 1$. The eigenvectors corresponding to the eigenvalues are:

$$\left\{ \begin{bmatrix} 10 \\ -2 \\ 3 \end{bmatrix} \right\} \leftrightarrow -1, \left\{ \begin{bmatrix} 7 \\ -2 \\ 2 \end{bmatrix} \right\} \leftrightarrow 1$$

Therefore this matrix is not diagonalizable.

6.2.2 The eigenvectors and eigenvalues are:

$$\left\{ \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\} \leftrightarrow 1, \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\} \leftrightarrow 1, \left\{ \begin{bmatrix} 7 \\ -2 \\ 2 \end{bmatrix} \right\} \leftrightarrow 3$$

The matrix P needed to diagonalize the above matrix is

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

and the diagonal matrix D is

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

6.2.3 The eigenvectors and eigenvalues are:

$$\left\{ \begin{bmatrix} -6 \\ -1 \\ -2 \end{bmatrix} \right\} \leftrightarrow 6, \left\{ \begin{bmatrix} -5 \\ -2 \\ 2 \end{bmatrix} \right\} \leftrightarrow -3, \left\{ \begin{bmatrix} -8 \\ -2 \\ 3 \end{bmatrix} \right\} \leftrightarrow -2$$

The matrix P needed to diagonalize the above matrix is

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

and the diagonal matrix D is

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

6.2.8 The eigenvalues are distinct because they are the n^{th} roots of 1. Hence if X is a given vector with

$$X = \sum_{j=1}^{n} a_j V_j$$

then

$$A^{nm} X = A^{nm} \sum_{j=1}^{n} a_j V_j = \sum_{j=1}^{n} a_j A^{nm} V_j = \sum_{j=1}^{n} a_j V_j = X$$

so $A^{nm} = I$.

6.3.1 First we write $A = PDP^{-1}$.

$$\left[\begin{array}{cc} 1 & 2 \\ 2 & 1 \end{array}\right] = \left[\begin{array}{rr} -1 & 1 \\ 1 & 1 \end{array}\right] \left[\begin{array}{rr} -1 & 0 \\ 0 & 3 \end{array}\right] \left[\begin{array}{rr} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{array}\right]$$

Therefore $A^{10} = PD^{10}P^{-1}$.

$$\begin{aligned}
\left[\begin{array}{cc} 1 & 2 \\ 2 & 1 \end{array}\right]^{10} &= \left[\begin{array}{rr} -1 & 1 \\ 1 & 1 \end{array}\right] \left[\begin{array}{rr} -1 & 0 \\ 0 & 3 \end{array}\right]^{10} \left[\begin{array}{rr} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{array}\right] \\
&= \left[\begin{array}{rr} -1 & 1 \\ 1 & 1 \end{array}\right] \left[\begin{array}{rr} (-1)^{10} & 0 \\ 0 & 3^{10} \end{array}\right] \left[\begin{array}{rr} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{array}\right] \\
&= \left[\begin{array}{cc} 29525 & 29524 \\ 29524 & 29525 \end{array}\right]
\end{aligned}$$

6.3.4

1. Multiply the given matrix by the initial state vector given by $\left[\begin{array}{c} 90 \\ 81 \\ 85 \end{array}\right]$. After one time period there are 89 people in location 1, 106 in location 2, and 61 in location 3.

2. Solve the system given by $(I - A)X_s = 0$ where A is the migration matrix and $X_s = \left[\begin{array}{c} x_{1s} \\ x_{2s} \\ x_{3s} \end{array}\right]$ is the steady state vector. The solution to this system is given by

$$\begin{aligned}
x_{1s} &= \frac{8}{5} x_{3s} \\
x_{2s} &= \frac{63}{25} x_{3s}
\end{aligned}$$

   Letting $x_{3s} = t$ and using the fact that there are a total of 256 individuals, we must solve

$$\frac{8}{5} t + \frac{63}{25} t + t = 256$$

   We find that $t = 50$. Therefore after a long time, there are 80 people in location 1, 126 in location 2, and 50 in location 3.

6.3.6 We solve $(I-A)X_s = 0$ to find the steady state vector $X_s = \begin{bmatrix} x_{1s} \\ x_{2s} \\ x_{3s} \end{bmatrix}$. The solution to the system is given by

$$\begin{aligned} x_{1s} &= \frac{5}{6}x_{3s} \\ x_{2s} &= \frac{2}{3}x_{3s} \end{aligned}$$

Letting $x_{3s} = t$ and using the fact that there are a total of 480 individuals, we must solve

$$\frac{5}{6}t + \frac{2}{3}t + t = 480$$

We find that $t = 192$. Therefore after a long time, there are 160 people in location 1, 128 in location 2, and 192 in location 3.

6.3.9 The migration matrix is

$$A = \begin{bmatrix} \frac{1}{3} & \frac{1}{10} & \frac{1}{10} & \frac{1}{5} \\ \frac{1}{3} & \frac{7}{10} & \frac{1}{5} & \frac{1}{10} \\ \frac{2}{9} & \frac{1}{10} & \frac{3}{5} & \frac{1}{5} \\ \frac{1}{9} & \frac{1}{10} & \frac{1}{10} & \frac{1}{2} \end{bmatrix}$$

To find the number of trailers in each location after a long time we solve system $(I-A)X_s = 0$ for the steady state vector $X_s = \begin{bmatrix} x_{1s} \\ x_{2s} \\ x_{3s} \\ x_{4s} \end{bmatrix}$. The solution to the system is

$$\begin{aligned} x_{1s} &= \frac{9}{10}x_{4s} \\ x_{2s} &= \frac{12}{5}x_{4s} \\ x_{3s} &= \frac{8}{5}x_{4s} \end{aligned}$$

Letting $x_{4s} = t$ and using the fact that there are a total of 413 trailers we must solve

$$\frac{9}{10}t + \frac{12}{5}t + \frac{8}{5}t + t = 413$$

We find that $t = 70$. Therefore after a long time, there are 63 trailers in the SE, 168 in the NE, 112 in the NW and 70 in the SW.

Index