MAT 04C LINEAR ALGEBRA



Zoya Kravets
Mission College



MAT 04C Linear Algebra (Kravets)

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide and accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

This text was compiled on 08/27/2025

TABLE OF CONTENTS

Licensing

1: Systems of Linear Equations

- 1.1: Systems of Linear Equations
- 1.2: Gaussian Elimination
- 1.3: Uniqueness of the Reduced Row-Echelon Form
- 1.4: Vectors
- 1.5: Vector Equations and Spans
- 1.6: Matrix Equations
- 1.7: Solution Sets
- 1.8: Applications of Linear Systems
- 1.9: Linear Independence
- 1.10: Exercises

2: Matrices

- 2.1: Matrix Arithmetic
- 2.2: Multiplication of Matrices
- 2.3: The ijth Entry of a Product
- 2.4: Properties of Matrix Multiplication
- 2.5: The Transpose
- 2.6: The Identity and Inverses
- 2.7: Finding the Inverse of a Matrix
- 2.8: Elementary Matrices
- 2.9: More on Matrix Inverses
- 2.10: LU Factorization
- 2.E: Exercises

3: Determinants

- 3.1: Basic Techniques
- 3.2: Properties of Determinants
- 3.3: Finding Determinants using Row Operations
- 3.4: Applications of the Determinant
- 3.E: Exercises

4: Linear Transformations

- 4.1: Linear Transformations
- 4.2: The Matrix of a Linear Transformation I
- 4.3: Properties of Linear Transformations. Invertible Linear Transformations.
- 4.4: Special Linear Transformations and Applications to Computer Graphics
- 4.5: One-to-One and Onto Transformations
- 4.6: Isomorphisms
- 4.7: The Kernel and Image of A Linear Map
- 4.8: The Matrix of a Linear Transformation II
- 4.9: The General Solution of a Linear System
- 4.E: Exercises

5: Vector Spaces and Subspaces

- 5.1: Vector Spaces
- 5.2: Subspaces
- 5.3: Basis and Dimension
- 5.4: Rank and Homogeneous Systems
- 5.5: The Rank Theorem
- 5.6: Bases as Coordinate Systems
- 5.7: Spanning, Linear Independence and Basis in R^n
- 5.8: A Linear Transformations
- 5.9: A Isomorphisms
- 5.10: The Kernel and Image of a Linear Map
- 5.11: The Matrix of a Linear Transformation
- 5.E: Exercises

6: Eigenvalues and Eigenvectors

- 6.1: Eigenvalues and Eigenvectors
- 6.2: The Characteristic Polynomial
- 6.3: Diagonalization
- 6.4: Complex Eigenvalues
- 6.5: Stochastic Matrices
- 6.6: Similarity

7: Orthogonality

- 7.1: Dot Products and Orthogonality
- 7.2: Orthogonal Complements
- 7.3: Orthogonal Projection
- 7.4: Orthogonal Sets and Gram-Schmidt Process
- 7.5: The Method of Least Squares
- 7.6: Orthogonal Diagonalization

8: Some Prerequisite Topics

- 8.1: Sets and Set Notation
- 8.2: Well Ordering and Mathematical Induction

Index

Glossary

Detailed Licensing

Detailed Licensing

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

1: Systems of Linear Equations

1.1: Systems of Linear Equations
1.2: Gaussian Elimination
1.3: Uniqueness of the Reduced Row-Echelon Form
1.4: Vectors
1.5: Vector Equations and Spans
1.6: Matrix Equations
1.7: Solution Sets
1.8: Applications of Linear Systems
1.9: Linear Independence
1.10: Exercises

This page titled 1: Systems of Linear Equations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

1.1: Systems of Linear Equations

(I) Systems of Linear 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.

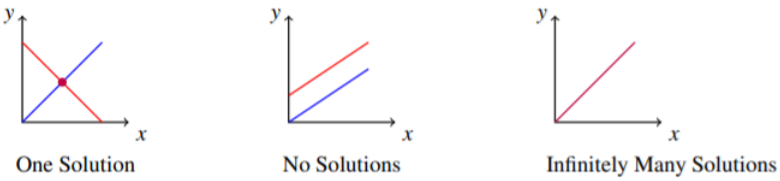


Figure 1.1.1

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 1.1.1: 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, $x=-1, y=4$.

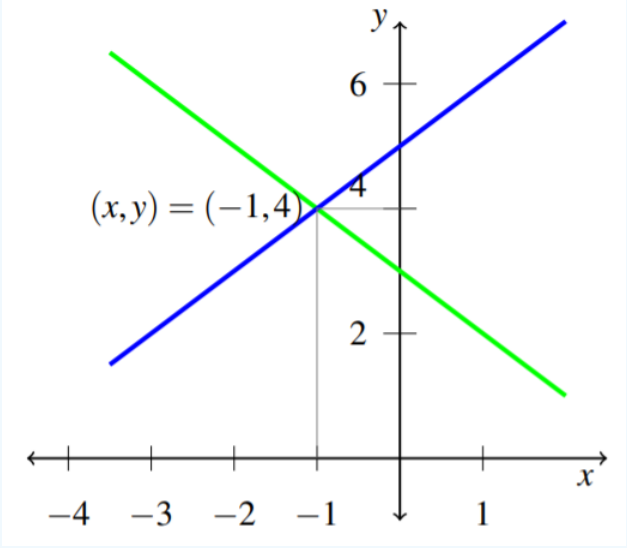


Figure 1.1.2

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.

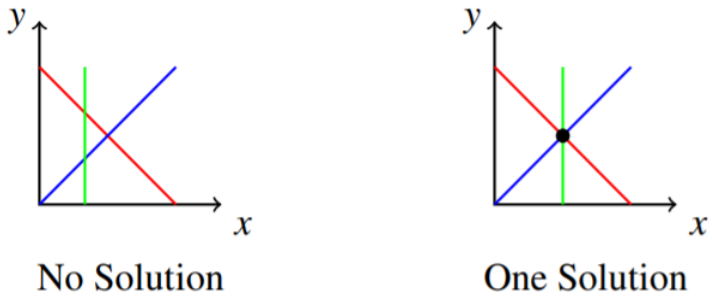


Figure 1.1.3

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.

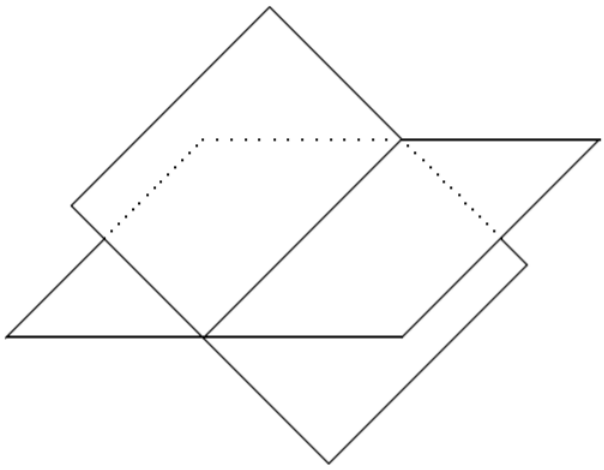


Figure 1.1.4

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.

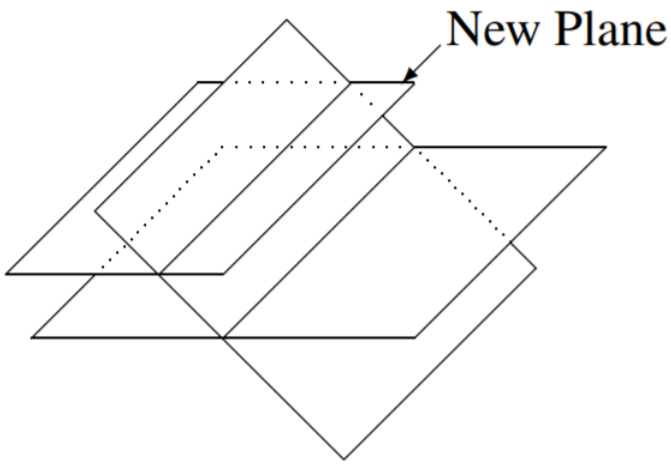


Figure 1.1.5

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.

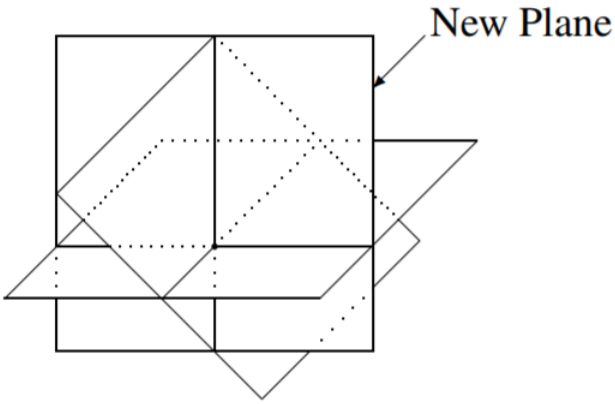


Figure 1.1.6

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.

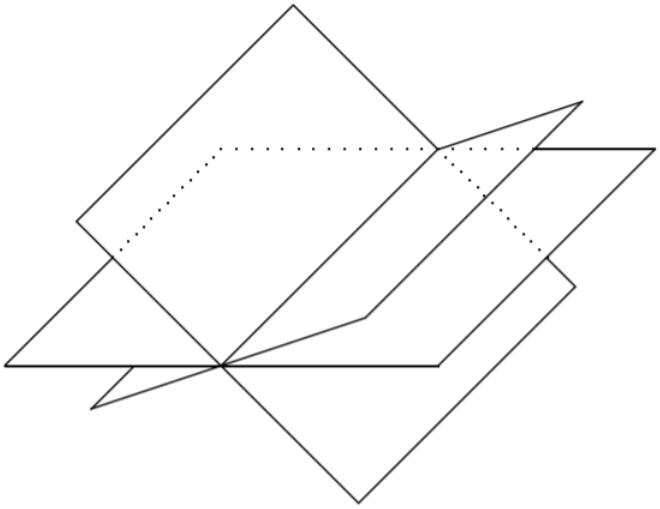


Figure 1.1.7

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.

? 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

a. No solution.
b. A unique solution.
c. An infinite number of solutions.

(II) Systems of Linear Equations, Algebra

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 .
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 1.1.1: System of Linear Equations

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

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

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 1.1.2: 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

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

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 1.1.3: 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.

Elementary Operations

We begin this section with an example. Recall from Example 1.1.1 that the solution to the given system was $(x, y) = (-1, 4)$.

Example 1.1.1: 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 1.1.4: 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 1.1.2: 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 1.1.4 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 1.1.1: Elementary Operations and Solutions

Suppose you have a system of two linear equations

$$\begin{aligned} E_1 &= b_1 \\ E_2 &= b_2 \end{aligned} \tag{1.1.1}$$

Then the following systems have the same solution set as (1.1.1):

1. $$\begin{aligned} E_2 &= b_2 \\ E_1 &= b_1 \end{aligned} \tag{1.1.2}$$

2. $$\begin{aligned} E_1 &= b_1 \\ kE_2 &= kb_2 \end{aligned} \tag{1.1.3}$$

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

3. $$\begin{aligned} E_1 &= b_1 \\ E_2 + kE_1 &= b_2 + kb_1 \end{aligned} \tag{1.1.4}$$

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

Before we proceed with the proof of Theorem 1.1.1, let us consider this theorem in context of Example 1.1.2. 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 1.1.1, 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 1.1.2, 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 1.1.1 tells us that the solution we found is in fact a solution to the original system.

We will now prove Theorem 1.1.1.

Proof

1. The proof that the systems (1.1.1) and (1.1.2) 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 (1.1.2) above. This is clear, because the system in (1.1.2) 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 (1.1.2).
2. Next we want to prove that the systems (1.1.1) and (1.1.3) 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 (1.1.3).
   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 (1.1.1) and (1.1.4) 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 (1.1.4). Then, we will show that any solution of (1.1.4) 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 (1.1.4). Similarly, it also solves $E_2 = b_2$. By our proof of (1.1.3), 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 (1.1.3), 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 1.1.3: Solving a System of Equations with Elementary Operations

Find the solutions to the system,

$$\begin{aligned} x + 3y + 6z &= 25 \\ 2x + 7y + 14z &= 58 \\ 2y + 5z &= 19 \end{aligned} \tag{1.1.5}$$

Solution

We can relate this system to Theorem 1.1.1 above. In this case, we have

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

Theorem 1.1.1 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 1.1.4. First, replace the second equation by (-2) times the first equation added to the second. This yields the system

$$\begin{aligned} x + 3y + 6z &= 25 \\ y + 2z &= 8 \\ 2y + 5z &= 19 \end{aligned} \tag{1.1.6}$$

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

$$\begin{aligned} x + 3y + 6z &= 25 \\ y + 2z &= 8 \\ z &= 3 \end{aligned} \tag{1.1.7}$$

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 (1.1.7) 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

$$x+3y=7$$
$$y=2$$
$$z=3$$

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

$$x=1$$
$$y=2$$
$$z=3$$

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, $(x,y,z)=(1,2,3)$.

This page titled 1.1: Systems of Linear Equations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.1: Systems of Equations, Geometry** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.
- **1.2: Systems of Equations, Algebraic Procedures** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

1.2: 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 1.2.3

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

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

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

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, $\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$. Similarly, we create a column from the coefficients on y in each equation, $\begin{bmatrix} 3 \\ 7 \\ 2 \end{bmatrix}$ and a column from the coefficients on z in each equation, $\begin{bmatrix} 6 \\ 14 \\ 5 \end{bmatrix}$. 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, $\begin{bmatrix} 25 \\ 58 \\ 19 \end{bmatrix}$.

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 1.2.1: Augmented Matrix of a Linear System

For a linear system of the form

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

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 1.2.4 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 1.2.1 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 1.2.2: 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 1.2.3. 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

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

Thus the first step in solving the system given by (1.2.5) would be to take (-2) times the first row of the augmented matrix and add it to the second row,

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

Note how this corresponds to (1.2.6). Next take (-2) times the second row and add to the third,

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

This augmented matrix corresponds to the system

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

which is the same as (1.2.7). 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 1.2.3: 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 1.2.4: 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 1.2.1: Not in Row-Echelon Form

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

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

Example 1.2.2: 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 1.2.3: 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 1.2.5: Pivot Position and Pivot Column

A **pivot position** in a matrix is the location of a leading entry in the row-echelon form of a matrix.

A **pivot column** is a column that contains a pivot position.

For example consider the following.

Example 1.2.4: 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:

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

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 1.2.1: 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 1.2.1.

Example 1.2.5: 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 1.2.1.

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

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

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

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

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 1.2.6: Finding the Solution to a System

Give the complete solution to the following system of equations

$$\begin{aligned} 2x + 4y - 3z &= -1 \\ 5x + 10y - 7z &= -2 \\ 3x + 6y + 5z &= 9 \end{aligned}$$

Solution

The augmented matrix for this system is

$$\left[\begin{array}{ccc|c} 2 & 4 & -3 & -1 \\ 5 & 10 & -7 & -2 \\ 3 & 6 & 5 & 9 \end{array}\right]$$

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 1.2.1. 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

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

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

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

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

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

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 1.2.7: An Infinite Set of Solutions

Give the complete solution to the system of equations

$$\begin{aligned} 3x-y-5z&=9 \\ y-10z&=0 \\ -2x+y&=-6 \end{aligned} \tag{1.2.1}$$

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 1.2.1. 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.

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

This is in reduced row-echelon form, which you should verify using Definition 1.2.4. 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 1.2.7 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 1.2.8: 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

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

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 1.2.1.

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

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

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

$$\left[\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 3 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{1.2.2}$$

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

$$\left[\begin{array}{c} x \\ y \\ z \\ w \end{array}\right] = \left[\begin{array}{c} -1+s-t \\ 2 \\ s \\ t \end{array}\right] \tag{1.2.3}$$

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 (1.2.2). 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 (1.2.3).

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.

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

or

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

This page titled 1.2: Gaussian Elimination is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.3: Gaussian Elimination** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

1.3: 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 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 1.3.8.

Example 1.3.1: Basic and Free Variables

Find the basic and free variables in the system

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

Solution

Recall from the solution of Example 1.3.8 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 1.3.1: 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 1.3.1: 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. 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 1.3.1 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 1.3.1: Equivalent Matrices

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

Proof

The proof of this theorem is left as an exercise.

Now, we can use Lemma 1.3.1 and Theorem 1.3.1 to prove the main result of this section.

Theorem 1.3.2: 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 1.3.1 ensures that all three homogeneous linear systems have exactly the same solutions. By Lemma 1.3.1 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.

This page titled 1.3: Uniqueness of the Reduced Row-Echelon Form is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.4: Uniqueness of the Reduced Row-Echelon Form** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

1.4: Vectors

Objectives

1. Learn how to add and scale vectors in $\mathbb{R}^n$, both algebraically and geometrically.
2. Understand linear combinations geometrically.
3. *Pictures:* vector addition, vector subtraction, linear combinations.
4. *Vocabulary words:* ***vector***, ***linear combination***.

Vectors in $\mathbb{R}^n$

We have been drawing points in $\mathbb{R}^n$ as dots in the line, plane, space, etc. We can also draw them as *arrows*. Since we have two geometric interpretations in mind, we now discuss the relationship between the two points of view.

Definition 1.4.1: Points and Vectors

Again, a ***point*** in $\mathbb{R}^n$ is drawn as a dot.

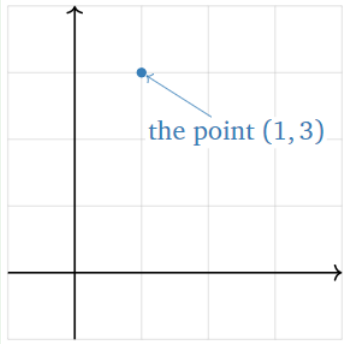


Figure 1.4.1

A ***vector*** is a point in $\mathbb{R}^n$, drawn as an arrow.

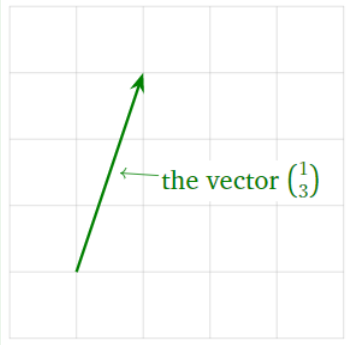


Figure 1.4.2

The difference is purely psychological: *points and vectors are both just lists of numbers*.

Example 1.4.1: Interactive: A vector in $\mathbb{R}^3$, by coordinates

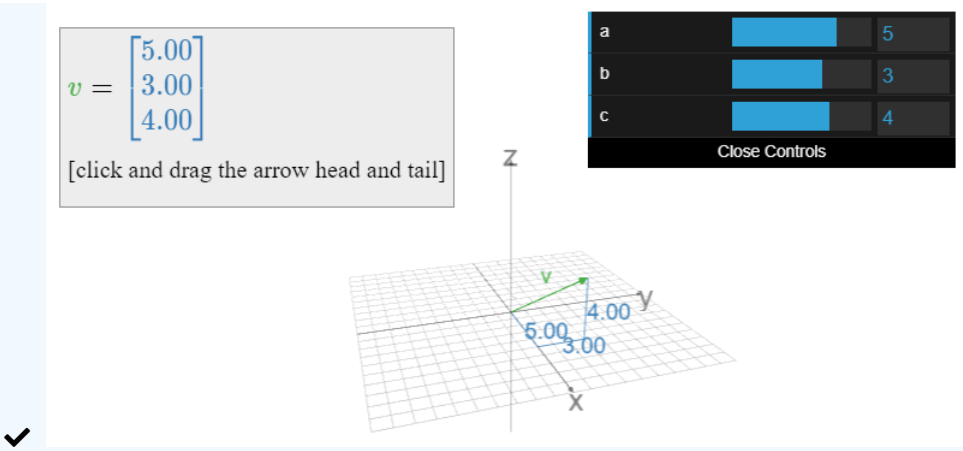


Figure 1.4.3. A vector in $\mathbb{R}^3$, and its coordinates. Drag the arrow head and tail.

When we think of a point in $\mathbb{R}^n$ as a vector, we will usually write it vertically, like a matrix with one column:

$$v = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

We will also write 0 for the zero vector.

Why make the distinction between points and vectors? A vector need not start at the origin: *it can be located anywhere*! In other words, an arrow is determined by its length and its direction, not by its location. For instance, these arrows all represent the vector $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

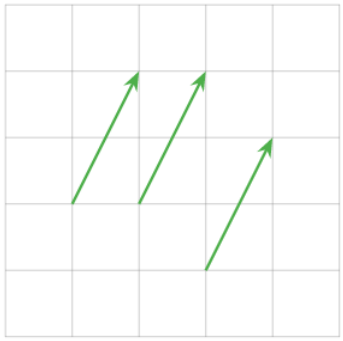


Figure 1.4.4

Note 1.4.1

Unless otherwise specified, we will assume that all vectors start at the origin.

Vectors makes sense in the real world: many physical quantities, such as velocity, are represented as vectors. But it makes more sense to think of the velocity of a car as being located at the car.

Remark

Some authors use boldface letters to represent vectors, as in "$\mathbf{v}$", or use arrows, as in "$\vec{v}$". As it is usually clear from context if a letter represents a vector, we do not decorate vectors in this way.

Note 1.4.2

Another way to think about a vector is as a *difference* between two points, or the arrow from one point to another. For instance, $\binom{1}{2}$ is the arrow from $(1,1)$ to $(2,3)$.

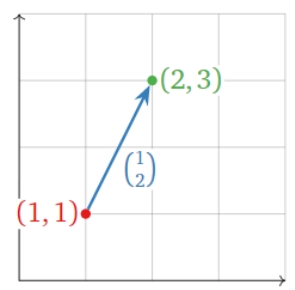


Figure 1.4.5

Vector Algebra and Geometry

Here we learn how to add vectors together and how to multiply vectors by numbers, both algebraically and geometrically.

Definition 1.4.2: Vector addition and scalar multiplication

- We can add two vectors together:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} + \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a+x \\ b+y \\ c+z \end{pmatrix}.$$

- We can multiply, or ***scale***, a vector by a real number c:

$$c\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} c \cdot x \\ c \cdot y \\ c \cdot z \end{pmatrix}.$$

We call c a ***scalar*** to distinguish it from a vector. If v is a vector and c is a scalar, then cv is called a **scalar multiple** of v.

Addition and scalar multiplication work in the same way for vectors of length n.

Example 1.4.2

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \\ 9 \end{pmatrix} \quad \text{and} \quad -2\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -2 \\ -4 \\ -6 \end{pmatrix}.$$

The Parallelogram Law for Vector Addition

Geometrically, the sum of two vectors v, w is obtained as follows: place the tail of w at the head of v. Then $v+w$ is the vector whose tail is the tail of v and whose head is the head of w. Doing this both ways creates a parallelogram. For example,

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

Why? The width of $v+w$ is the sum of the widths, and likewise with the heights.

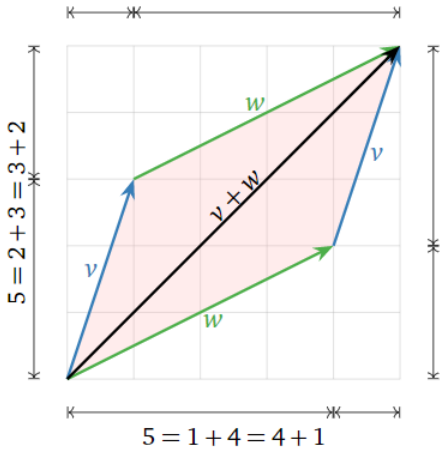


Figure 1.4.6

Example 1.4.3: Interactive: The parallelogram law for vector addition

$$\begin{bmatrix} 3.00 \\ -5.00 \\ 4.00 \end{bmatrix} + \begin{bmatrix} 4.00 \\ -1.00 \\ -2.00 \end{bmatrix} = \begin{bmatrix} 7.00 \\ -6.00 \\ 2.00 \end{bmatrix}$$

[click and drag the heads of v and w to move them]

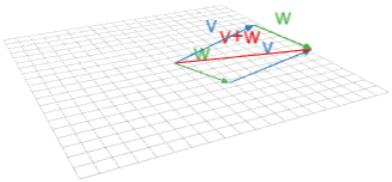


Figure 1.4.7: The parallelogram law for vector addition. Click and drag the heads of and w.

Vector Subtraction

Geometrically, the difference of two vectors v, w is obtained as follows: place the tail of v and w at the same point. Then $v - w$ is the vector from the head of w to the head of v. For example,

$$\begin{pmatrix} 1 \\ 4 \end{pmatrix} - \begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} -3 \\ 2 \end{pmatrix}.$$

Why? If you add $v - w$ to w, you get v.

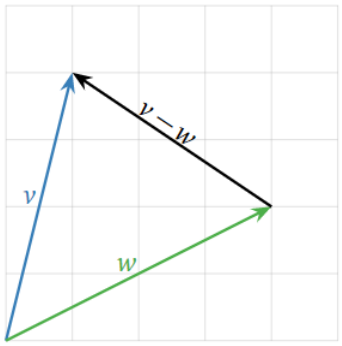


Figure 1.4.8

Example 1.4.4: Interactive: Vector subtraction

$$\begin{bmatrix} 3.00 \\ -5.00 \\ 4.00 \end{bmatrix} - \begin{bmatrix} 4.00 \\ -1.00 \\ -2.00 \end{bmatrix} = \begin{bmatrix} -1.00 \\ -4.00 \\ 6.00 \end{bmatrix}$$

[click and drag the heads of v and w to move them]

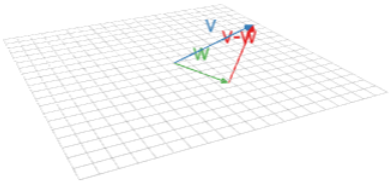


Figure 1.4.9: Vector subtraction. Click and drag the heads of v and w.

Scalar Multiplication

A scalar multiple of a vector v has the same (or opposite) direction, but a different length. For instance, $2v$ is the vector in the direction of v but twice as long, and $-\frac{1}{2}v$ is the vector in the opposite direction of v, but half as long. Note that the set of all scalar multiples of a (nonzero) vector v is a *line*.

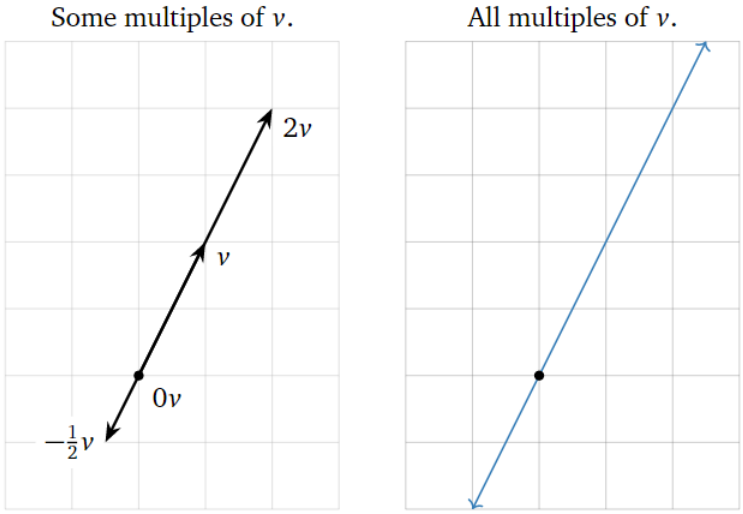


Figure 1.4.10

Example 1.4.5: Interactive: Scalar multiplication

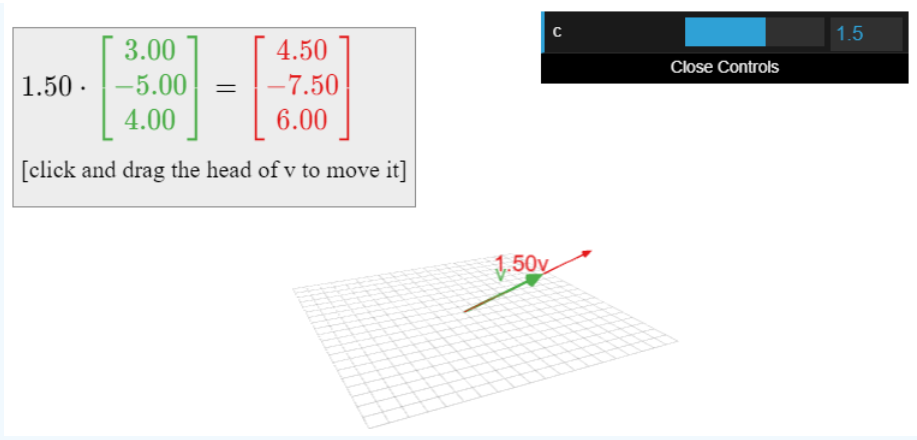


Figure 1.4.11: Scalar multiplication. Drag the slider to change the scalar.

Linear Combinations

We can add and scale vectors in the same equation.

Definition 1.4.3: Linear Combination

Let $c_1, c_2, \ldots, c_k$ be scalars, and let $v_1, v_2, \ldots, v_k$ be vectors in $\mathbb{R}^n$. The vector in $\mathbb{R}^n$

$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

is called a ***linear combination*** of the vectors $v_1, v_2, \ldots, v_k$, with ***weights*** or ***coefficients*** $c_1, c_2, \ldots, c_k$.

Geometrically, a linear combination is obtained by stretching / shrinking the vectors $v_1, v_2, \ldots, v_k$ according to the coefficients, then adding them together using the parallelogram law.

Example 1.4.6

Let $v_1 = \binom{1}{2}$ and $v_2 = \binom{1}{0}$. Here are some linear combinations of v_1 and v_2, drawn as points.

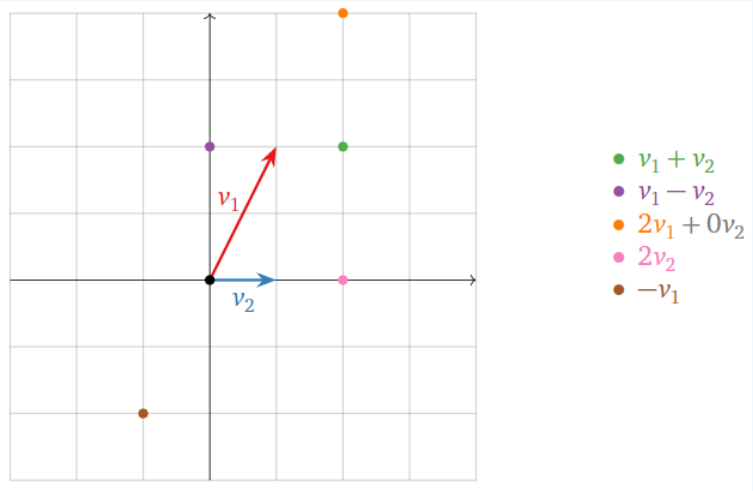


Figure 1.4.12

The locations of these points are found using the parallelogram law for vector addition. Any vector on the plane is a linear combination of v_1 and v_2, with suitable coefficients.

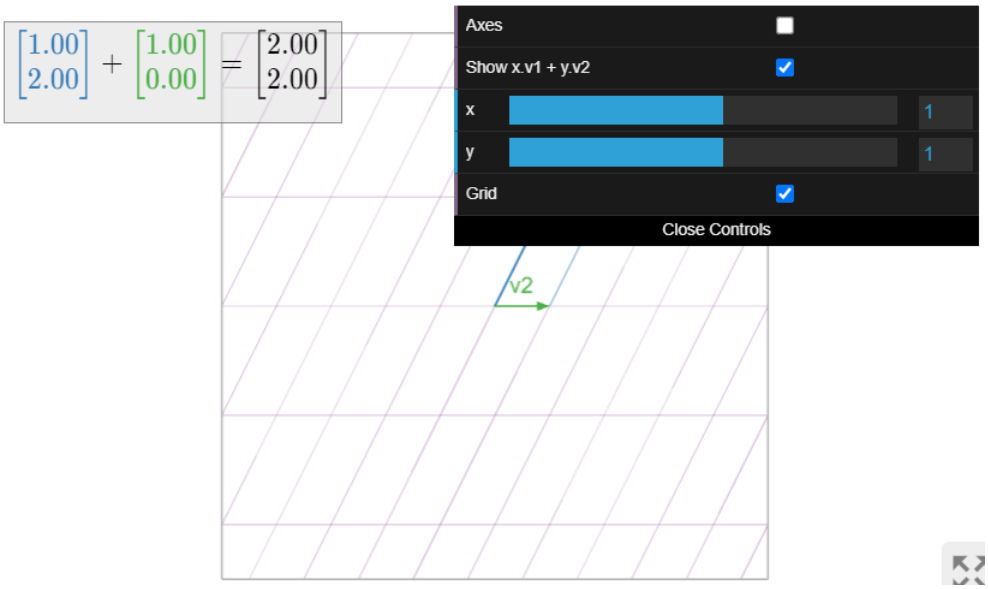


Figure 1.4.13: Linear combinations of two vectors in $\mathbb{R}^2$: move the sliders to change the coefficients of v_1 and v_2. Note that any vector on the plane can be obtained as a linear combination of v_1, v_2 with suitable coefficients.

Example 1.4.7: Interactive: Linear combinations of three vectors

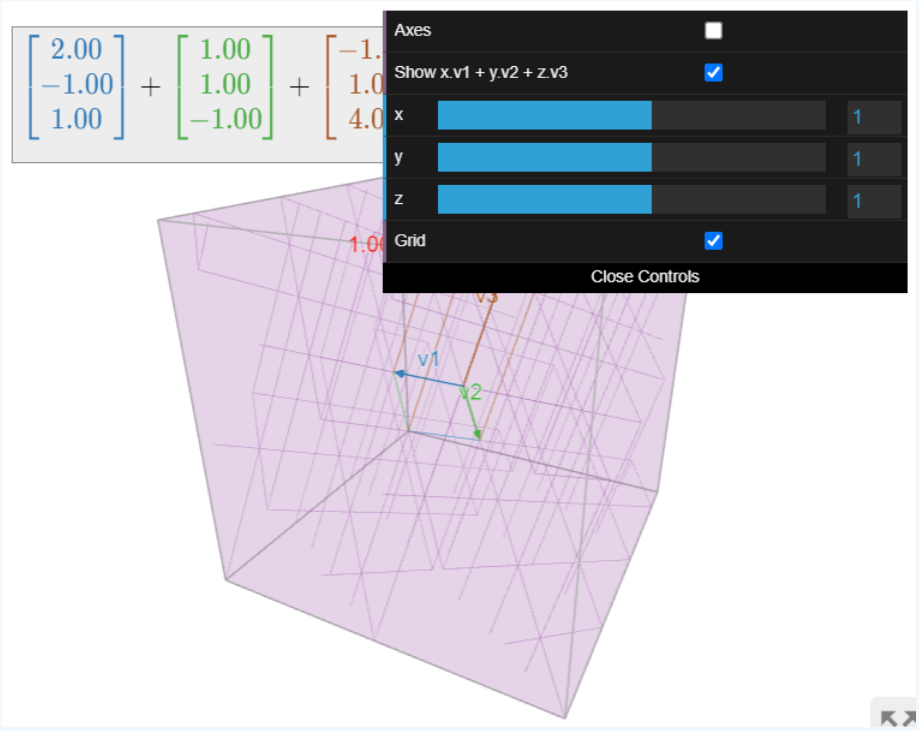


Figure 1.4.14: Linear combinations of three vectors: move the sliders to change the coefficients of v_1, v_2, v_3. Note how the parallelogram law for addition of three vectors is more of a "parallepiped law".

Example 1.4.8: Linear Combinations of a Single Vector

A linear combination of a single vector $v = \binom{1}{2}$ is just a scalar multiple of v. So some examples include

$$v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \frac{3}{2}v = \begin{pmatrix} 3/2 \\ 3 \end{pmatrix}, \quad -\frac{1}{2}v = \begin{pmatrix} -1/2 \\ -1 \end{pmatrix}, \quad \cdots$$

The set of all linear combinations is the *line through* v. (Unless $v = 0$, in which case any scalar multiple of v is again 0.)

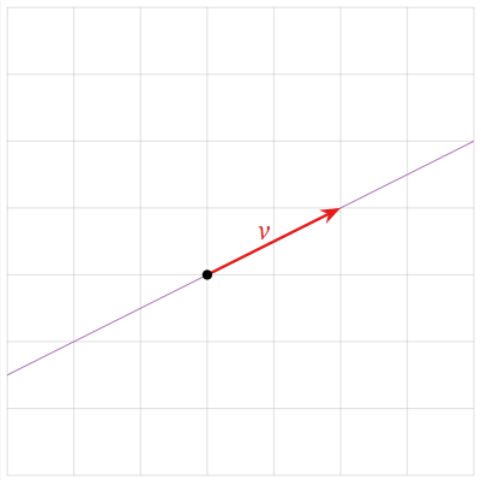


Figure 1.4.15

Example 1.4.9: Linear Combinations of Collinear Vectors

The set of all linear combinations of the vectors

$$v_1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

is the *line* containing both vectors.

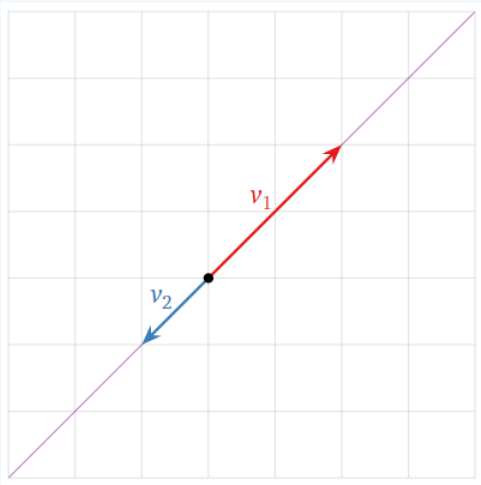


Figure 1.4.16

The difference between this and Example 1.4.6 is that both vectors lie on the same line. Hence any scalar multiples of v_1, v_2 lie on that line, as does their sum.

Example 1.4.10: Interactive: Linear combinations of two collinear vectors

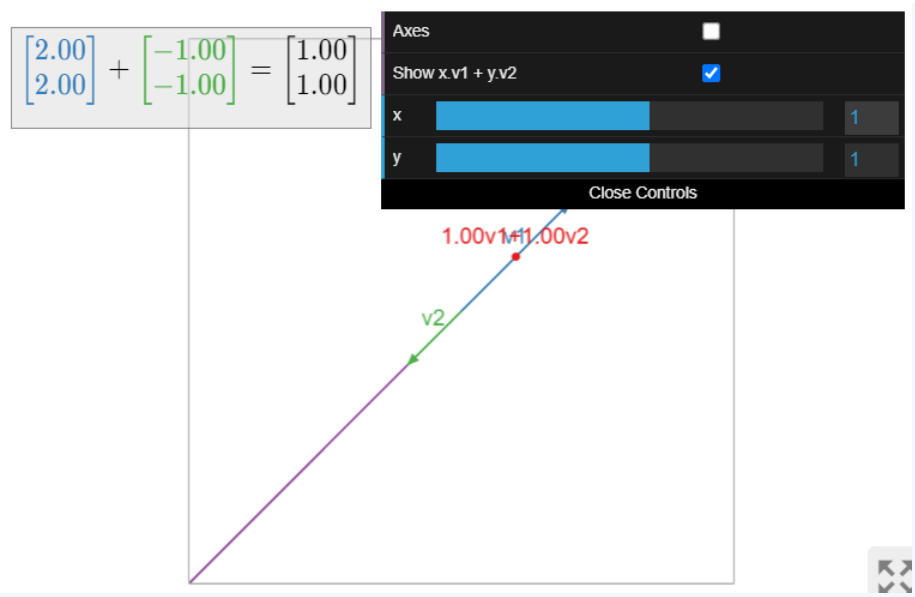


Figure 1.4.17: Linear combinations of two collinear vectors in $\mathbb{R}^2$. Move the sliders to change the coefficients of v_1, v_2. Note that there is no way to "escape" the line.

This page titled 1.4: Vectors is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **2.1: Vectors** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

1.5: Vector Equations and Spans

Objectives

1. Understand the equivalence between a system of linear equations and a vector equation.
2. Learn the definition of $\text{Span}\{x_1, x_2, \ldots, x_k\}$, and how to draw pictures of spans.
3. *Recipe:* solve a vector equation using augmented matrices / decide if a vector is in a span.
4. *Pictures:* an inconsistent system of equations, a consistent system of equations, spans in $\mathbb{R}^2$ and $\mathbb{R}^3$.
5. *Vocabulary word:* ***vector equation***.
6. *Essential vocabulary word:* ***span***.

Vector Equations

An equation involving vectors with n coordinates is the same as n equations involving only numbers. For example, the equation

$$x\begin{pmatrix}1\\2\\6\end{pmatrix} + y\begin{pmatrix}-1\\-2\\-1\end{pmatrix} = \begin{pmatrix}8\\16\\3\end{pmatrix}$$

simplifies to

$$\begin{pmatrix}x\\2x\\6x\end{pmatrix} + \begin{pmatrix}-y\\-2y\\-y\end{pmatrix} = \begin{pmatrix}8\\16\\3\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}x-y\\2x-2y\\6x-y\end{pmatrix} = \begin{pmatrix}8\\16\\3\end{pmatrix}.$$

For two vectors to be equal, all of their coordinates must be equal, so this is just the system of linear equations

$$\begin{cases} x - y = 8 \\ 2x - 2y = 16 \\ 6x - y = 3. \end{cases}$$

Definition 1.5.1: Vector Equation

A ***vector equation*** is an equation involving a linear combination of vectors with possibly unknown coefficients.

Note 1.5.1

Asking whether or not a vector equation has a solution is the same as asking if a given vector is a linear combination of some other given vectors.

For example the vector equation above is asking if the vector $(8, 16, 3)$ is a linear combination of the vectors $(1, 2, 6)$ and $(-1, 2, -1)$.

The thing we really care about is solving systems of linear equations, not solving vector equations. The whole point of vector equations is that they give us a different, and more geometric, way of viewing systems of linear equations.

Note 1.5.2: A Picture of a Consistent System

Below we will show that the above system of equations is consistent. Equivalently, this means that the above vector equation has a solution. In other words, there is a linear combination of $(1, 2, 6)$ and $(-1, 2, -1)$ that equals $(8, 16, 3)$. We can visualize the last statement geometrically. Therefore, the following Figure 1.5.1 gives a *picture of a consistent system of equations.* Compare with Figure 1.5.2, which shows a picture of an inconsistent system.

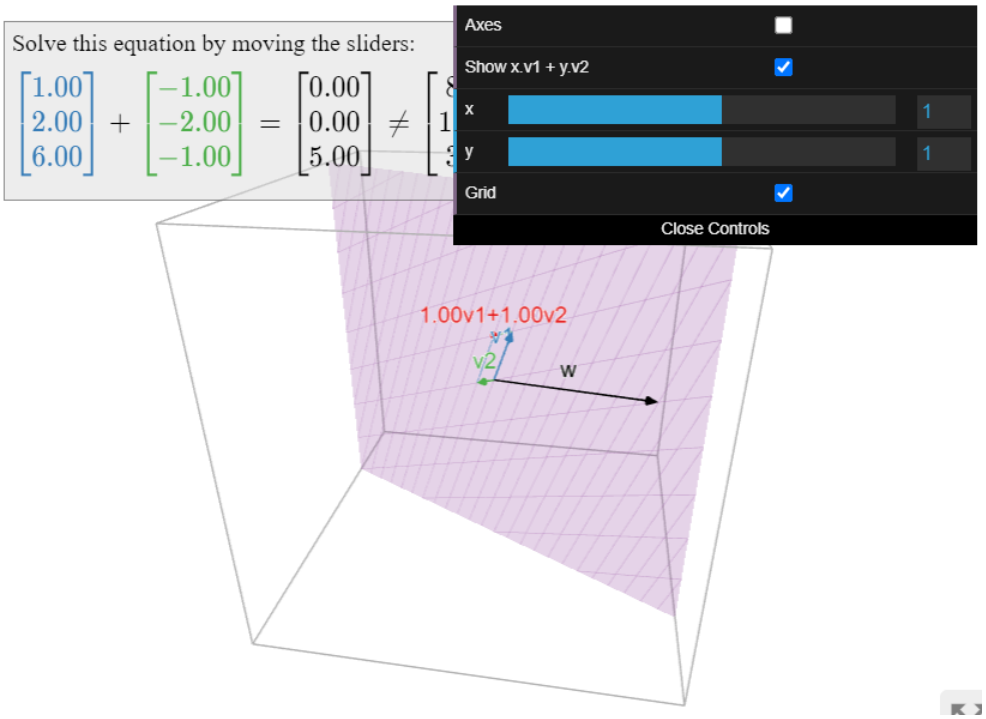


Figure 1.5.1: A picture of the above vector equation. Try to solve the equation geometrically by moving the sliders.

In order to actually solve the vector equation

$$x\begin{pmatrix}1\\2\\6\end{pmatrix}+y\begin{pmatrix}-1\\-2\\-1\end{pmatrix}=\begin{pmatrix}8\\16\\3\end{pmatrix},$$

one has to solve the system of linear equations

$$\begin{cases} x - y = 8 \\ 2x - 2y = 16 \\ 6x - y = 3. \end{cases}$$

This means forming the augmented matrix

$$\left(\begin{array}{cc|c} 1 & -1 & 8 \\ 2 & -2 & 16 \\ 6 & -1 & 3 \end{array}\right)$$

and row reducing. Note that *the columns of the augmented matrix are the vectors from the original vector equation*, so it is not actually necessary to write the system of equations: one can go directly from the vector equation to the augmented matrix by "smooshing the vectors together". In Example 1.5.1 we carry out the row reduction and find the solution.

Example 1.5.1

Is $\begin{pmatrix}8\\16\\3\end{pmatrix}$ a linear combination of $\begin{pmatrix}1\\2\\6\end{pmatrix}$ and $\begin{pmatrix}-1\\-2\\-1\end{pmatrix}$?

Solution

As discussed above, this question boils down to a row reduction:

$$\left(\begin{array}{cc|c} 1 & -1 & 8 \\ 2 & -2 & 16 \\ 6 & -1 & 3 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & -9 \\ 0 & 0 & 0 \end{array}\right).$$

From this we see that the equation is consistent, and the solution is $x = -1$ and $y = -9$. We conclude that $\begin{pmatrix} 8 \\ 16 \\ 3 \end{pmatrix}$ is indeed a linear combination of $\begin{pmatrix} 1 \\ 2 \\ 6 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ -2 \\ -1 \end{pmatrix}$, with coefficients -1 and -9:

$$-\begin{pmatrix} 1 \\ 2 \\ 6 \end{pmatrix} - 9\begin{pmatrix} -1 \\ -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 8 \\ 6 \\ 3 \end{pmatrix}.$$

Recipe: Solving a Vector Equation

In general, the vector equation

$$x_1v_1 + x_2v_2 + \cdots + x_kv_k = b$$

where $v_1, v_2, \ldots, v_k$, b are vectors in $\mathbb{R}^n$ and $x_1, x_2, \ldots, x_k$ are unknown scalars, has the same solution set as the linear system with augmented matrix

$$\left(\begin{array}{cccc|c} | & | & & | & | \\ v_1 & v_2 & \cdots & v_k & b \\ | & | & & | & | \end{array}\right)$$

whose columns are the v_i's and the b's.

Now we have *three* equivalent ways of thinking about a linear system:

1. As a system of equations:

$$\begin{cases} 2x_1 + 3x_2 - 2x_3 = 7 \\ x_1 - x_2 - 3x_3 = 5 \end{cases}$$

2. As an augmented matrix:

$$\left(\begin{array}{ccc|c} 2 & 3 & -2 & 7 \\ 1 & -1 & -3 & 5 \end{array}\right)$$

3. As a vector equation ($x_1v_1 + x_2v_2 + \cdots + x_nv_n = b$):

$$x_1\begin{pmatrix} 2 \\ 1 \end{pmatrix} + x_2\begin{pmatrix} 3 \\ -1 \end{pmatrix} + x_3\begin{pmatrix} -2 \\ -3 \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \end{pmatrix}$$

The third is geometric in nature: it lends itself to drawing pictures.

Spans

It will be important to know what are *all* linear combinations of a set of vectors $v_1, v_2, \ldots, v_k$ in $\mathbb{R}^n$. In other words, we would like to understand the set of all vectors b in $\mathbb{R}^n$ such that the vector equation (in the unknowns $x_1, x_2, \ldots, x_k$)

$$x_1v_1 + x_2v_2 + \cdots + x_kv_k = b$$

has a solution (i.e. is consistent).

Definition 1.5.2: Span

Let $v_1, v_2, \ldots, v_k$ be vectors in $\mathbb{R}^n$. The ***span*** of $v_1, v_2, \ldots, v_k$ is the collection of all linear combinations of $v_1, v_2, \ldots, v_k$, and is denoted $\text{Span}\{v_1, v_2, \ldots, v_k\}$. In symbols:

$$\text{Span}\{v_1, v_2, \ldots, v_k\} = \{x_1v_1 + x_2v_2 + \cdots + x_kv_k \mid x_1, x_2, \ldots, x_k \text{ in } \mathbb{R}\}$$

We also say that $\text{Span}\{v_1, v_2, \ldots, v_k\}$ is the subset ***spanned by*** or ***generated by*** the vectors $v_1, v_2, \ldots, v_k$.

The above definition, Definition 1.5.2 is the first of several *essential definitions* that we will see in this textbook. They are essential in that they form the essence of the subject of linear algebra: learning linear algebra means (in part) learning these definitions. All of the definitions are important, but it is essential that you learn and understand the definitions marked as such.

Note 1.5.3: Set Builder Notation

The notation

$$\{x_1v_1 + x_2v_2 + \cdots + x_kv_k \mid x_1, x_2, \ldots, x_k \text{ in } \mathbb{R}\}$$

reads as: "the set of all things of the form $x_1v_1 + x_2v_2 + \cdots + x_kv_k$ such that $x_1, x_2, \ldots, x_k$ are in $\mathbb{R}$." The vertical line is "such that"; everything to the left of it is "the set of all things of this form", and everything to the right is the condition that those things must satisfy to be in the set. Specifying a set in this way is called ***set builder notation***.

All mathematical notation is only shorthand: any sequence of symbols must translate into a usual sentence.

Note 1.5.4: Three characterizations of consistency

Now we have three equivalent ways of making the same statement:

1. A vector b is in the span of $v_1, v_2, \ldots, v_k$.
2. The vector equation

$$x_1v_1 + x_2v_2 + \cdots + x_kv_k = b$$

has a solution.

3. The linear system with augmented matrix

$$\left(\begin{array}{cccc|c} | & | & & | & | \\ v_1 & v_2 & \cdots & v_k & b \\ | & | & & | & | \end{array}\right)$$

is consistent.

Equivalent means that, for any given list of vectors $v_1, v_2, \ldots, v_k,\ b$, either all three statements are true, or all three statements are false.

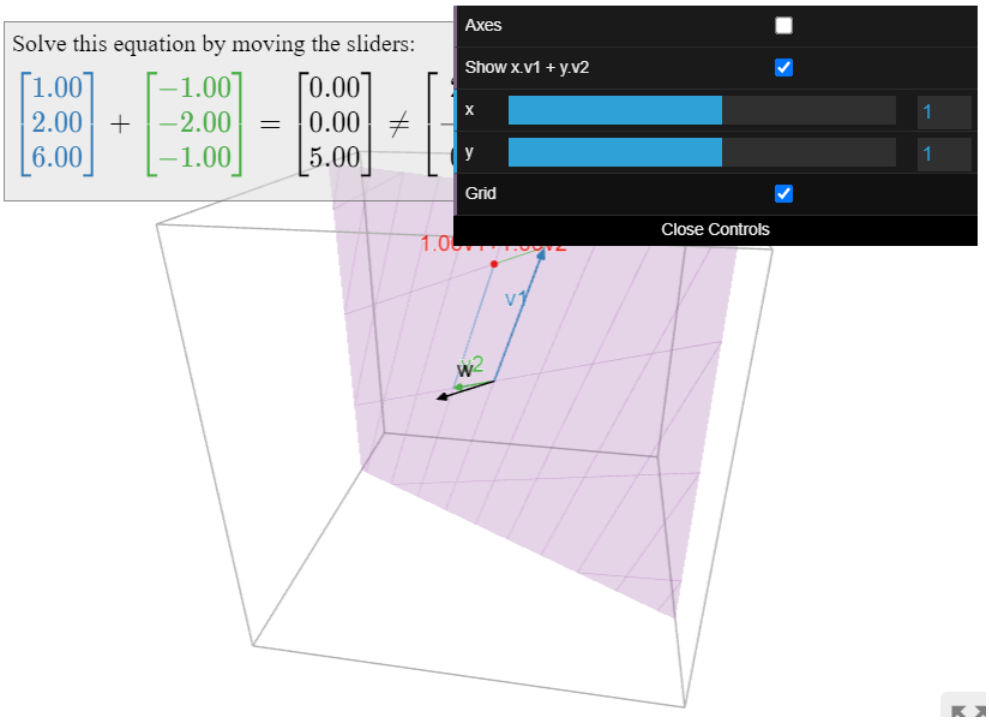


Figure 1.5.2: This is a picture of an *inconsistent* linear system: the vector w on the right-hand side of the equation $x_1v_1 + x_2v_2 = w$ is not in the span of v_1, v_2. Convince yourself of this by trying to solve the equation $x_1v_1 + x_2v_2 = w$ by moving the sliders, and by row reduction. Compare this with Figure 1.5.1.

Pictures of Spans.

Drawing a picture of $\text{Span}\{v_1, v_2, \ldots, v_k\}$ is the same as drawing a picture of all linear combinations of $v_1, v_2, \ldots, v_k$.

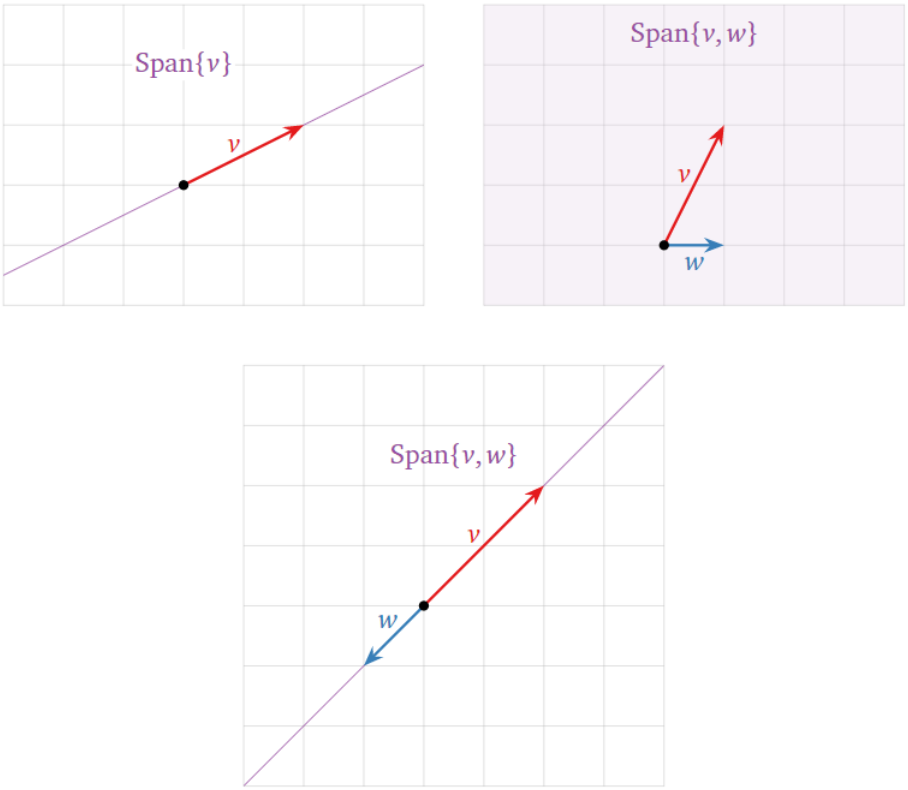


Figure 1.5.3: Pictures of spans in $\mathbb{R}^2$.

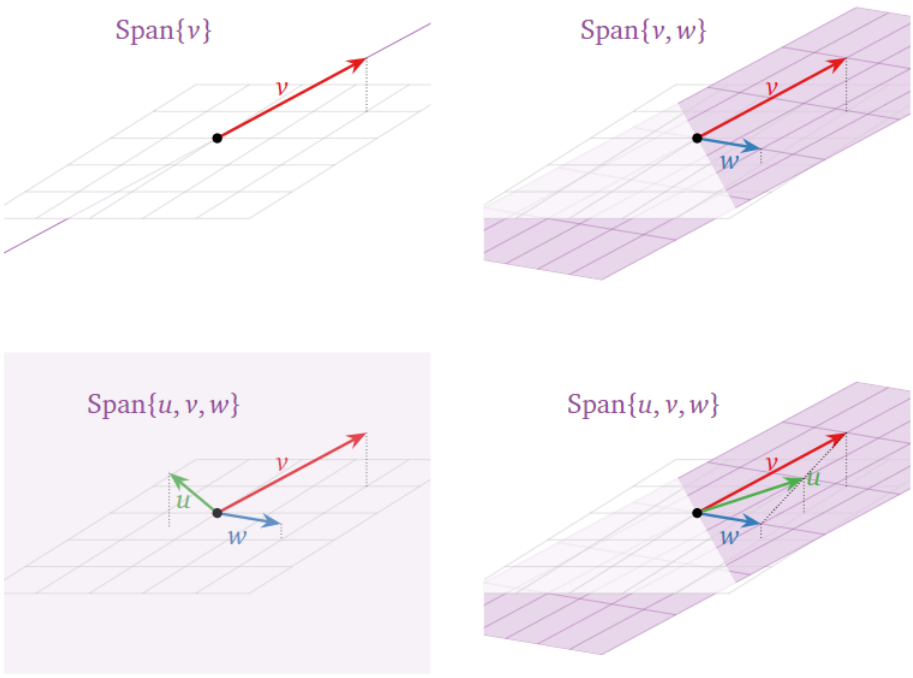


Figure 1.5.4: Pictures of spans in $\mathbb{R}^3$. The span of two noncollinear vectors is the plane containing the origin and the heads of the vectors. Note that three coplanar (but not collinear) vectors span a plane and not a 3-space, just as two collinear vectors span a line and not a plane.

Example 1.5.2: Interactive: Span of two vectors in $\mathbb{R}^2$

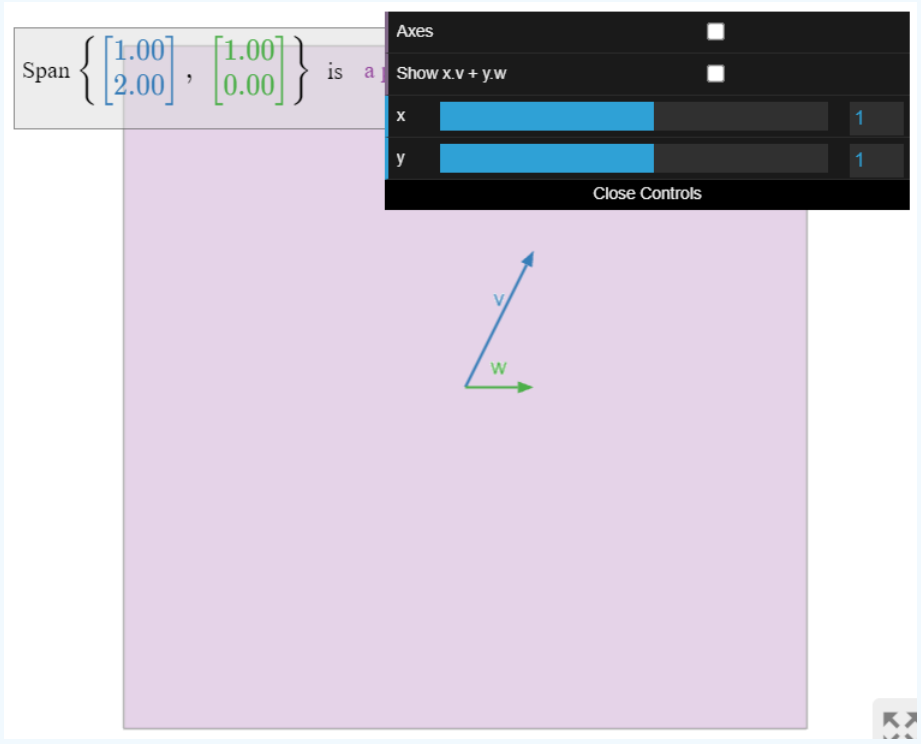


Figure 1.5.5: Interactive picture of a span of two vectors in $\mathbb{R}^2$. Check "Show x.v + y.w" and move the sliders to see how every point in the violet region is in fact a linear combination of the two vectors.

Example 1.5.3: Interactive: Span of two vectors in $\mathbb{R}^3$

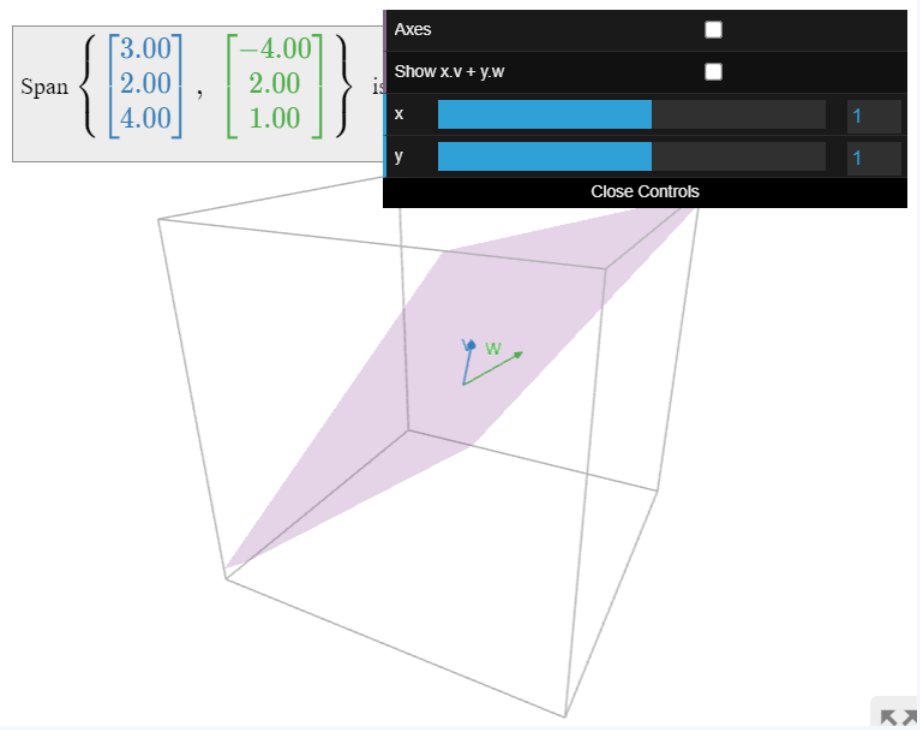


Figure 1.5.6: Interactive picture of a span of two vectors in $\mathbb{R}^3$. Check "Show x.v + y.w" and move the sliders to see how every point in the violet region is in fact a linear combination of the two vectors.

Example 1.5.4: Interactive: Span of three vectors in $\mathbb{R}^3$

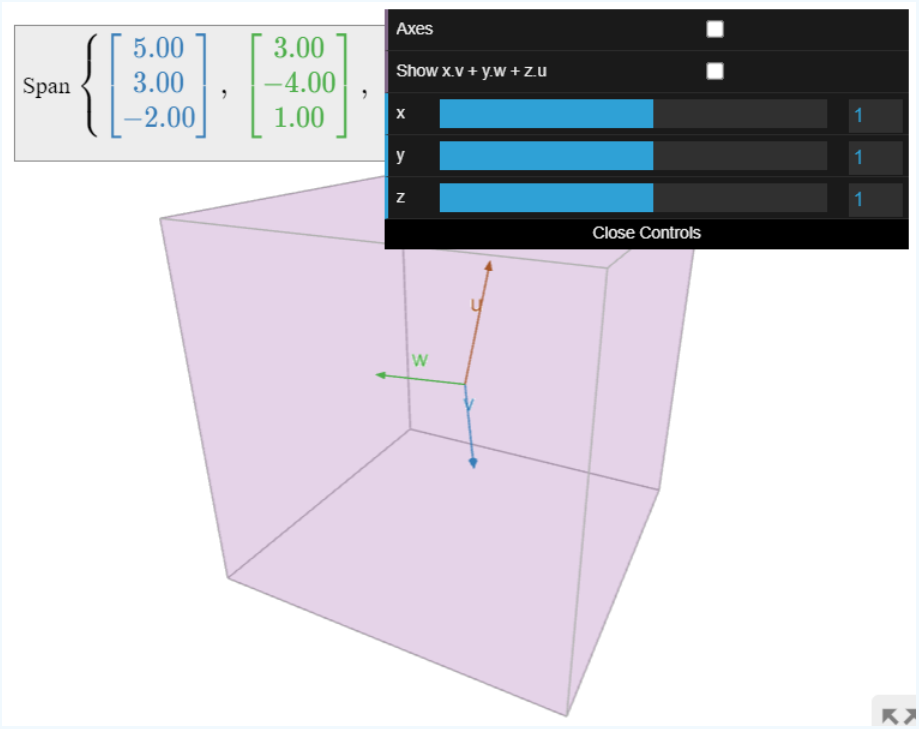


Figure 1.5.7: Interactive picture of a span of three vectors in $\mathbb{R}^3$. Check "Show x.v + y.w + z.u" and move the sliders to see how every point in the violet region is in fact a linear combination of the three vectors.

This page titled 1.5: Vector Equations and Spans is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **2.2: Vector Equations and Spans** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

1.6: Matrix Equations

Objectives

1. Understand the equivalence between a system of linear equations, an augmented matrix, a vector equation, and a matrix equation.
2. Characterize the vectors b such that $Ax = b$ is consistent, in terms of the span of the columns of A.
3. Characterize matrices A such that $Ax = b$ is consistent for all vectors b.
4. *Recipe:* multiply a vector by a matrix (two ways).
5. *Picture:* the set of all vectors b such that $Ax = b$ is consistent.
6. *Vocabulary word:* ***matrix equation***.

The Matrix Equation $Ax = b$

In this section we introduce a very concise way of writing a system of linear equations: $Ax = b$. Here A is a matrix and x, b are vectors (generally of different sizes), so first we must explain how to multiply a matrix by a vector.

Note 1.6.1

When we say "A is an $m \times n$ matrix," we mean that A has m rows and n columns.

Remark

In this book, we do *not* reserve the letters m and n for the numbers of rows and columns of a matrix. If we write "A is an $n \times m$ matrix", then n is the number of rows of A and m is the number of columns.

Definition 1.6.1: Product

Let A be an $m \times n$ matrix with columns $v_1, v_2, \ldots, v_n$:

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}$$

The ***product*** of A with a vector x in $\mathbb{R}^n$ is the linear combination

$$Ax = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 v_1 + x_2 v_2 + \cdots + x_n v_n.$$

This is a vector in $\mathbb{R}^m$.

Example 1.6.1

$$\begin{pmatrix} 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 1 \begin{pmatrix} 4 \\ 7 \end{pmatrix} + 2 \begin{pmatrix} 5 \\ 8 \end{pmatrix} + 3 \begin{pmatrix} 6 \\ 9 \end{pmatrix} = \begin{pmatrix} 32 \\ 50 \end{pmatrix}.$$

In order for Ax to make sense, the number of entries of x has to be the same as the number of columns of A: we are using the entries of x as the coefficients of the columns of A in a linear combination. The resulting vector has the same number of entries as the number of *rows* of A, since each column of A has that number of entries.

Note 1.6.2

If A is an $m \times n$ matrix (m rows, n columns), then Ax makes sense when x has n entries. The product Ax has m entries.

Proposition 1.6.1: Properties of the Matrix-Vector Product

Let A be an $m \times n$ matrix, let u, v be vectors in $\mathbb{R}^n$, and let c be a scalar. Then:

- $A(u+v) = Au + Av$
- $A(cu) = cAu$

Definition 1.6.2: Matrix Equation

A ***matrix equation*** is an equation of the form $Ax = b$, where A is an $m \times n$ matrix, b is a vector in $\mathbb{R}^m$, and x is a vector whose coefficients $x_1, x_2, \ldots, x_n$ are unknown.

In this book we will study two complementary questions about a matrix equation $Ax = b$:

1. Given a specific choice of b, what are all of the solutions to $Ax = b$?
2. What are all of the choices of b so that $Ax = b$ is consistent?

The first question is more like the questions you might be used to from your earlier courses in algebra; you have a lot of practice solving equations like $x^2 - 1 = 0$ for x. The second question is perhaps a new concept for you. Theorem 2.9.1 in Section 2.9, which is the culmination of this chapter, tells us that the two questions are intimately related.

Note 1.6.3: Matrix Equations and Vector Equations

Let $v_1, v_2, \ldots, v_n$ and b be vectors in $\mathbb{R}^m$. Consider the vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = b.$$

This is equivalent to the matrix equation $Ax = b$, where

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \quad \text{and} \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Conversely, if A is any $m \times n$ matrix, then $Ax = b$ is equivalent to the vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = b,$$

where $v_1, v_2, \ldots, v_n$ are the columns of A, and $x_1, x_2, \ldots, x_n$ are the entries of x.

Example 1.6.2

Write the vector equation

$$2v_1 + 3v_2 - 4v_3 = \begin{pmatrix} 7 \\ 2 \\ 1 \end{pmatrix}$$

as a matrix equation, where v_1, v_2, v_3 are vectors in $\mathbb{R}^3$.

Solution

Let A be the matrix with columns v_1, v_2, v_3, and let x be the vector with entries $2, 3, -4$. Then

$$Ax = \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ -4 \end{pmatrix} = 2v_1 + 3v_2 - 4v_3,$$

so the vector equation is equivalent to the matrix equation $Ax = \begin{pmatrix} 7 \\ 2 \\ 1 \end{pmatrix}$.

Note 1.6.4: Four Ways of Writing a Linear System

We now have *four* equivalent ways of writing (and thinking about) a system of linear equations:

1. As a system of equations:

$$\begin{cases} 2x_1 + 3x_2 - 2x_3 = 7 \\ x_1 - x_2 - 3x_3 = 5 \end{cases}$$

2. As an augmented matrix:

$$\left(\begin{array}{ccc|c} 2 & 3 & -2 & 7 \\ 1 & -1 & -3 & 5 \end{array}\right)$$

3. As a vector equation ($x_1v_1 + x_2v_2 + \cdots + x_nv_n = b$):

$$x_1 \begin{pmatrix} 2 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} 3 \\ -1 \end{pmatrix} + x_3 \begin{pmatrix} -2 \\ -3 \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \end{pmatrix}$$

4. As a matrix equation ($Ax = b$):

$$\begin{pmatrix} 2 & 3 & -2 \\ 1 & -1 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \end{pmatrix}.$$

In particular, *all four have the same solution set*.

Note 1.6.5

We will move back and forth freely between the four ways of writing a linear system, over and over again, for the rest of the book.

Another Way to Compute Ax

The above definition for Product, Definition 1.6.1, is a useful way of defining the product of a matrix with a vector when it comes to understanding the relationship between matrix equations and vector equations. Here we give a definition that is better-adapted to computations by hand.

Definition 1.6.3: Row Vector

A ***row vector*** is a matrix with one row. The ***product*** of a row vector of length n and a (column) vector of length n is

$$\begin{pmatrix} a_1 & a_2 & \cdots a_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = a_1x_1 + a_2x_2 + \cdots + a_nx_n.$$

This is a scalar.

Recipe: The Row-Column Rule for Matrix-Vector Multiplication

If A is an $m \times n$ matrix with rows $r_1, r_2, \ldots, r_m$, and x is a vector in $\mathbb{R}^n$, then

$$Ax = \begin{pmatrix} -r_1- \\ -r_2- \\ \vdots \\ -r_m- \end{pmatrix} x = \begin{pmatrix} r_1 x \\ r_2 x \\ \vdots \\ r_m x \end{pmatrix}.$$

Example 1.6.3

$$\begin{pmatrix} 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \\ \\ \begin{pmatrix} 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 4 \cdot 1 + 5 \cdot 2 + 6 \cdot 3 \\ 7 \cdot 1 + 8 \cdot 2 + 9 \cdot 3 \end{pmatrix} = \begin{pmatrix} 32 \\ 50 \end{pmatrix}.$$

This is the same answer as before:

$$\begin{pmatrix} 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 1 \begin{pmatrix} 4 \\ 7 \end{pmatrix} + 2 \begin{pmatrix} 5 \\ 8 \end{pmatrix} + 3 \begin{pmatrix} 6 \\ 9 \end{pmatrix} = \begin{pmatrix} 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 \\ 1 \cdot 7 + 2 \cdot 8 + 3 \cdot 9 \end{pmatrix} = \begin{pmatrix} 32 \\ 50 \end{pmatrix}.$$

Spans and Consistency

Let A be a matrix with columns $v_1, v_2, \ldots, v_n$:

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}.$$

Then

$$\begin{aligned} &Ax = b \text{ has a solution} \\ &\iff \text{there exist } x_1, x_2, \ldots, x_n \text{ such that } A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = b \\ &\iff \text{there exist } x_1, x_2, \ldots, x_n \text{ such that } x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = b \\ &\iff b \text{ is a linear combination of } v_1, v_2, \ldots, v_n \\ &\iff b \text{ is in the span of the columns of } A. \end{aligned}$$

Note 1.6.6: Spans and Consistency

The matrix equation $Ax = b$ has a solution if and only if b is in the span of the columns of A.

This gives an equivalence between an *algebraic* statement ($Ax = b$ is consistent), and a *geometric* statement (b is in the span of the columns of A).

Example 1.6.4: An Inconsistent System

Let $A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \\ 1 & -1 \end{pmatrix}$. Does the equation $Ax = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$ have a solution?

Solution

First we answer the question geometrically. The columns of A are

$$v_1 = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix},$$

and the target vector (on the right-hand side of the equation) is $w = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$. The equation $Ax = w$ is consistent if and only if w is contained in the span of the columns of A. So we draw a picture:

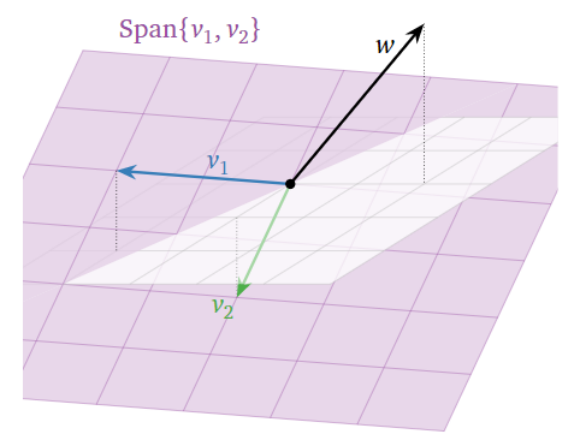


Figure 1.6.1

It does not appear that w lies in $\text{Span}\{v_1, v_2\}$, so the equation is inconsistent.

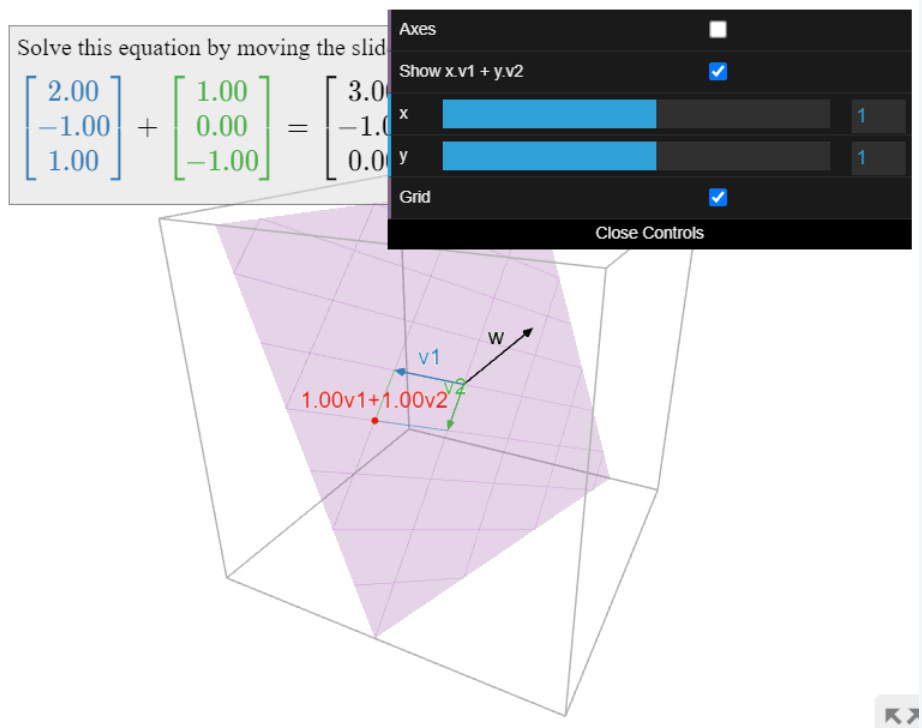


Figure 1.6.2: The vector w is not contained in $\text{Span}\{v_1, v_2\}$, so the equation $Ax = b$ is inconsistent. (Try moving the sliders to solve the equation.)

Let us check our geometric answer by solving the matrix equation using row reduction. We put the system into an augmented matrix and row reduce:

$$\left(\begin{array}{cc|c} 2 & 1 & 0 \\ -1 & 0 & 2 \\ 1 & -1 & 2 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right).$$

The last equation is $0 = 1$, so the system is indeed inconsistent, and the matrix equation

$$\begin{pmatrix} 2 & 1 \\ -1 & 0 \\ 1 & -1 \end{pmatrix} x = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$$

has no solution.

Example 1.6.5: A Consistent System

Let $A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \\ 1 & -1 \end{pmatrix}$. Does the equation $Ax = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$ have a solution?

Solution

First we answer the question geometrically. The columns of A are

$$v_1 = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix},$$

and the target vector (on the right-hand side of the equation) is $w = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$. The equation $Ax = w$ is consistent if and only if w is contained in the span of the columns of A. So we draw a picture:

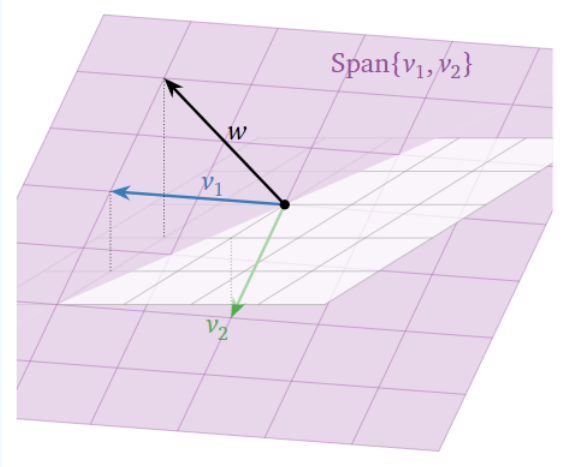


Figure 1.6.3

It appears that w is indeed contained in the span of the columns of A; in fact, we can see

$$w = v_1 - v_2 \implies x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

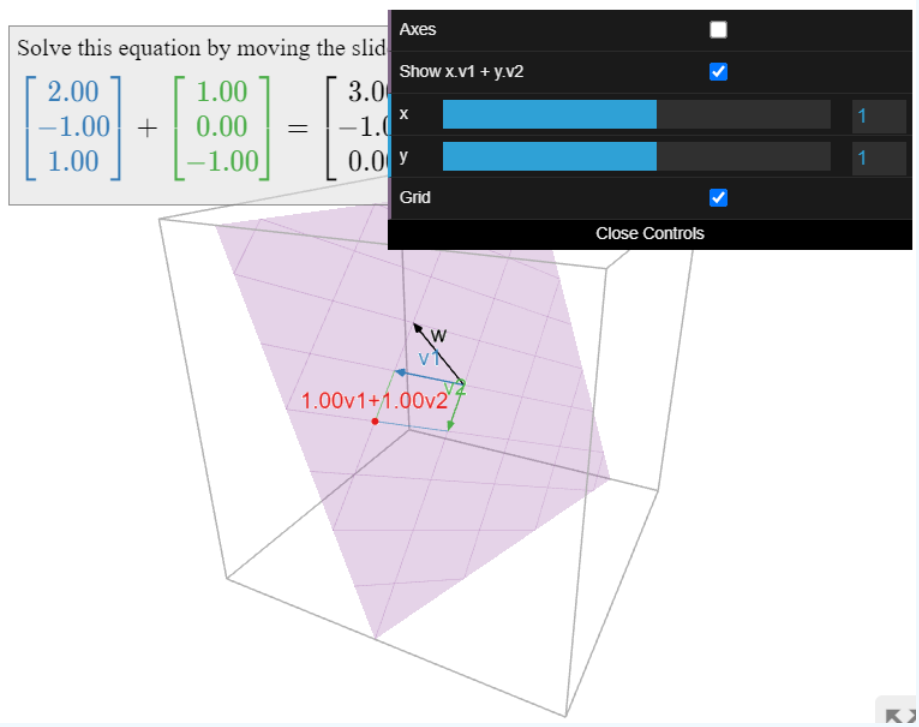


Figure 1.6.4: The vector w is contained in $\mathrm{Span}\{v_1, v_2\}$, so the equation $Ax = b$ is consistent. (Move the sliders to solve the equation.)

Let us check our geometric answer by solving the matrix equation using row reduction. We put the system into an augmented matrix and row reduce:

$$\left(\begin{array}{cc|c} 2 & 1 & 1 \\ -1 & 0 & -1 \\ 1 & -1 & 2 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{array}\right).$$

This gives us $x = 1$ and $y = -1$, which is consistent with the picture:

$$1\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} - 1\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \quad \text{or} \quad A\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}.$$

When Solutions Always Exist.

Building on the Note 1.6.6: Spans and Consistency, we have the following criterion for when $Ax = b$ is consistent for *every* choice of b.

Theorem 1.6.1

Let A be an $m \times n$ (non-augmented) matrix. The following are equivalent:

1. $Ax = b$ has a solution for all b in $\mathbb{R}^m$.
2. The span of the columns of A is all of $\mathbb{R}^m$.
3. A has a pivot position, Definition 1.2.5 in Section 1.2, in every row.

Proof

The equivalence of 1 and 2 is established by the Note 1.6.6 Spans and Consistency as applied to every b in $\mathbb{R}^m$.

Now we show that 1 and 3 are equivalent. (Since we know 1 and 2 are equivalent, this implies 2 and 3 are equivalent as well.) If A has a pivot in every row, then its reduced row echelon form looks like this:

$$\begin{pmatrix} 1 & 0 & \star & 0 & \star \\ 0 & 1 & \star & 0 & \star \\ 0 & 0 & 0 & 1 & \star \end{pmatrix},$$

and therefore $\left(\begin{array}{c|c} A & b \end{array}\right)$ reduces to this:

$$\left(\begin{array}{ccccc|c} 1 & 0 & \star & 0 & \star & \star \\ 0 & 1 & \star & 0 & \star & \star \\ 0 & 0 & 0 & 1 & \star & \star \end{array}\right).$$

There is no b that makes it inconsistent, so there is always a solution. Conversely, if A does not have a pivot in each row, then its reduced row echelon form looks like this:

$$\begin{pmatrix} 1 & 0 & \star & 0 & \star \\ 0 & 1 & \star & 0 & \star \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

which can give rise to an inconsistent system after augmenting with b:

$$\left(\begin{array}{ccccc|c} 1 & 0 & \star & 0 & \star & 0 \\ 0 & 1 & \star & 0 & \star & 0 \\ 0 & 0 & 0 & 0 & 0 & 16 \end{array}\right).$$

Recall that ***equivalent*** means that, for any given matrix A, either *all* of the conditions of the above Theorem, 1.6.1, are true, or they are all false.

Note 1.6.7

Be careful when reading the statement of the above Theorem, 1.6.1. The first two conditions look very much like this Note 1.6.6: Spans and Consistency, but they are logically quite different because of the quantifier "*for all* b".

Example 1.6.6: Interactive: The criteria of the theorem are satisfied

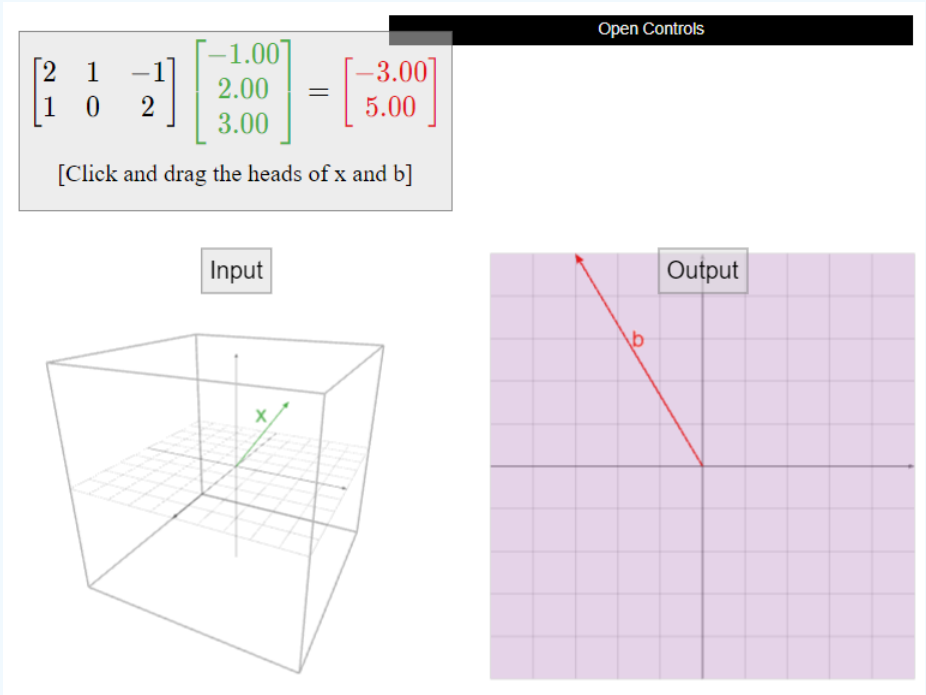


Figure 1.6.5: An example where the criteria of the above Theorem 1.6.1 are satisfied. The violet region is the span of the columns v_1, v_2, v_3 of A, which is the same as the set of all b such that $Ax = b$ has a solution. If you drag b, the demo will solve $Ax = b$ for you and move x.

Example 1.6.7: Interactive: The critera of the theorem are not satisfied

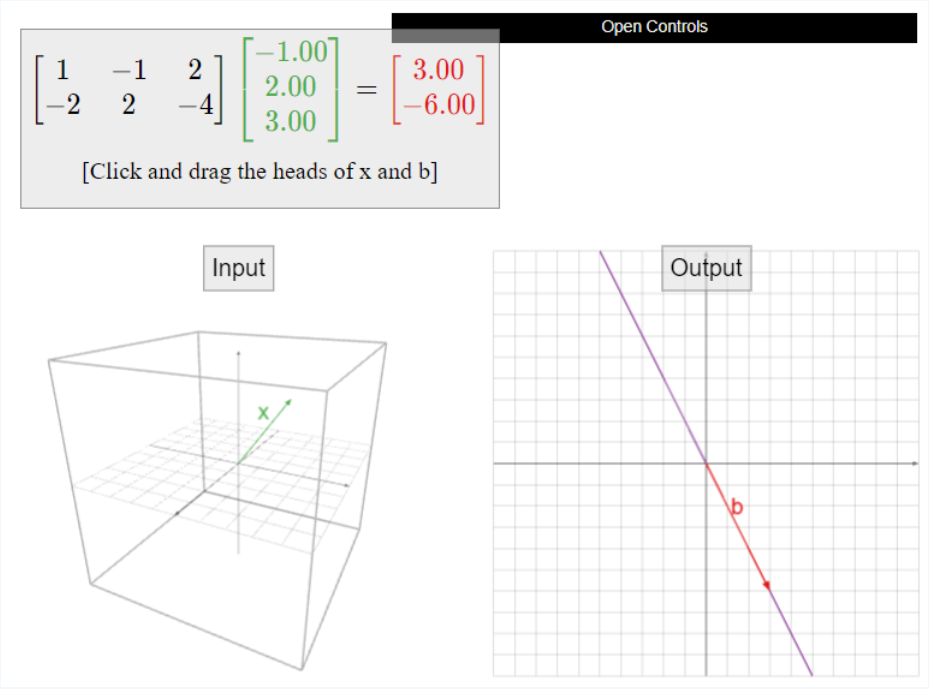


Figure 1.6.6: An example where the criteria of the above Theorem 1.6.1 are *not* satisfied. The violet line is the span of the columns v_1, v_2, v_3 of A, which is the same as the set of all b such that $Ax = b$ has a solution. Try dragging b in and out of the column span.

Remark

We will see in this Corollary 2.7.1 in Section 2.7 that the dimension of the span of the columns is equal to the number of pivots of A. That is, the columns of A span a line if A has one pivot, they span a plane if A has two pivots, etc. The whole space $\mathbb{R}^m$ has dimension m, so this generalizes the fact that the columns of A span $\mathbb{R}^m$ when A has m pivots.

This page titled 1.6: Matrix Equations is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **2.3: Matrix Equations** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

1.7: Solution Sets

Objectives

1. Understand the relationship between the solution set of $Ax = 0$ and the solution set of $Ax = b$.
2. Understand the difference between the solution set and the column span.
3. *Recipes:* parametric vector form, write the solution set of a homogeneous system as a span.
4. *Pictures:* solution set of a homogeneous system, solution set of an inhomogeneous system, the relationship between the two.
5. *Vocabulary words:* ***homogeneous*/*inhomogeneous***, ***trivial solution***.

In this section we will study the geometry of the solution set of any matrix equation $Ax = b$.

Homogeneous Systems

The equation $Ax = b$ is easier to solve when $b = 0$, so we start with this case.

Definition 1.7.1: Homogeneous & Inhomogeneous

A system of linear equations of the form $Ax = 0$ is called ***homogeneous***.

A system of linear equations of the form $Ax = b$ for $b \neq 0$ is called ***inhomogeneous***.

A homogeneous system is just a system of linear equations where all constants on the right side of the equals sign are zero.

Note 1.7.1

A homogeneous system always has the solution $x = 0$. This is called the ***trivial solution***. Any nonzero solution is called ***nontrivial***.

Observation 1.7.1

The equation $Ax = 0$ has a nontrivial solution $\iff$ there is a free variable $\iff$ A has a column without a pivot position.

Example 1.7.1: No nontrivial solutions

What is the solution set of $Ax = 0$, where

$$A = \begin{pmatrix} 1 & 3 & 4 \\ 2 & -1 & 2 \\ 1 & 0 & 1 \end{pmatrix}?$$

Solution

We form an augmented matrix and row reduce:

$$\left(\begin{array}{ccc|c} 1 & 3 & 4 & 0 \\ 2 & -1 & 2 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right).$$

The only solution is the trivial solution $x = 0$.

Observation 1.7.2

When we row reduce the augmented matrix for a homogeneous system of linear equations, the last column will be zero throughout the row reduction process. We saw this in the last Example 1.7.1:

$$\left(\begin{array}{ccc|c} 1 & 3 & 4 & 0 \\ 2 & -1 & 2 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right)$$

So it is not really necessary to write augmented matrices when solving homogeneous systems.

When the homogeneous equation $Ax = 0$ does have nontrivial solutions, it turns out that the solution set can be conveniently expressed as a span.

Example 1.7.2: Parametric Vector Form (homogeneous case)

Consider the following matrix in reduced row echelon form:

$$A = \begin{pmatrix} 1 & 0 & -8 & -7 \\ 0 & 1 & 4 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The matrix equation $Ax = 0$ corresponds to the system of equations

$$\begin{cases} x_1 - 8x_3 - 7x_4 = 0 \\ x_2 + 4x_3 + 3x_4 = 0. \end{cases}$$

We can write the parametric form as follows:

$$\begin{cases} x_1 = 8x_3 + 7x_4 \\ x_2 = -4x_3 - 3x_4 \\ x_3 = x_3 \\ x_4 = x_4. \end{cases}$$

We wrote the redundant equations $x_3 = x_3$ and $x_4 = x_4$ in order to turn the above system into a *vector equation*:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_3 \begin{pmatrix} 8 \\ -4 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 7 \\ -3 \\ 0 \\ 1 \end{pmatrix}.$$

This vector equation is called the ***parametric vector form*** of the solution set. Since x_3 and x_4 are allowed to be anything, this says that the solution set is the set of all linear combinations of $\begin{pmatrix} 8 \\ -4 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 7 \\ -3 \\ 0 \\ 1 \end{pmatrix}$. In other words, the solution set is

$$\text{Span}\left\{\begin{pmatrix} 8 \\ -4 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 7 \\ -3 \\ 0 \\ 1 \end{pmatrix}\right\}.$$

Here is the general procedure.

Recipe: Parametric Vector Form (Homogeneous Case)

Let A be an $m \times n$ matrix. Suppose that the free variables in the homogeneous equation $Ax = 0$ are, for example, x_3, x_6, and x_8.

1. Find the reduced row echelon form of A.
2. Write the parametric form of the solution set, including the redundant equations $x_3 = x_3$, $x_6 = x_6$, $x_8 = x_8$. Put equations for all of the x_i in order.

3. Make a single vector equation from these equations by making the coefficients of x_3, x_6, and x_8 into vectors v_3, v_6, and v_8, respectively.

The solutions to $Ax = 0$ will then be expressed in the form

$$x = x_3v_3 + x_6v_6 + x_8v_8$$

for some vectors v_3, v_6, v_8 in $\mathbb{R}^n$, and any scalars x_3, x_6, x_8. This is called the ***parametric vector form*** of the solution.

In this case, the solution set can be written as $\text{Span}\{v_3,\ v_6,\ v_8\}$.

We emphasize the following fact in particular.

Note 1.7.2

The set of solutions to a homogeneous equation $Ax = 0$ is a span.

Example 1.7.3: The solution set is a line

Compute the parametric vector form of the solution set of $Ax = 0$, where

$$A = \begin{pmatrix} 1 & -3 \\ 2 & -6 \end{pmatrix}.$$

Solution

We row reduce (without augmenting, as suggested in the above Observation 1.7.2):

$$\begin{pmatrix} 1 & -3 \\ 2 & -6 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & -3 \\ 0 & 0 \end{pmatrix}.$$

This corresponds to the single equation $x_1 - 3x_2 = 0$. We write the parametric form including the redundant equation $x_2 = x_2$:

$$\begin{cases} x_1 &= 3x_2 \\ x_2 &= x_2. \end{cases}$$

We turn these into a single vector equation:

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_2 \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

This is the parametric vector form of the solution set. Since x_2 is allowed to be anything, this says that the solution set is the set of all scalar multiples of $\binom{3}{1}$, otherwise known as

$$\text{Span}\left\{\begin{pmatrix} 3 \\ 1 \end{pmatrix}\right\}.$$

We know how to draw the picture of a span of a vector: it is a line. Therefore, this is a picture of the the solution set:

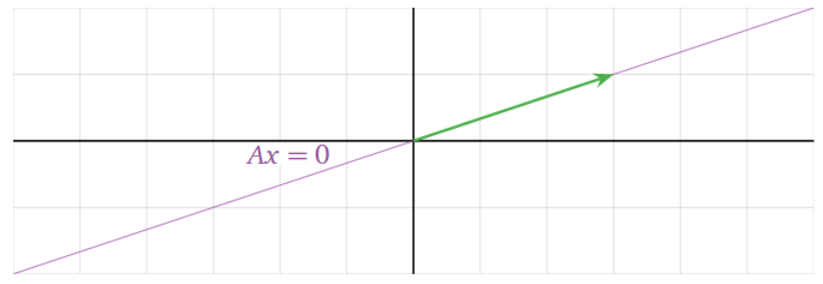


Figure 1.7.1

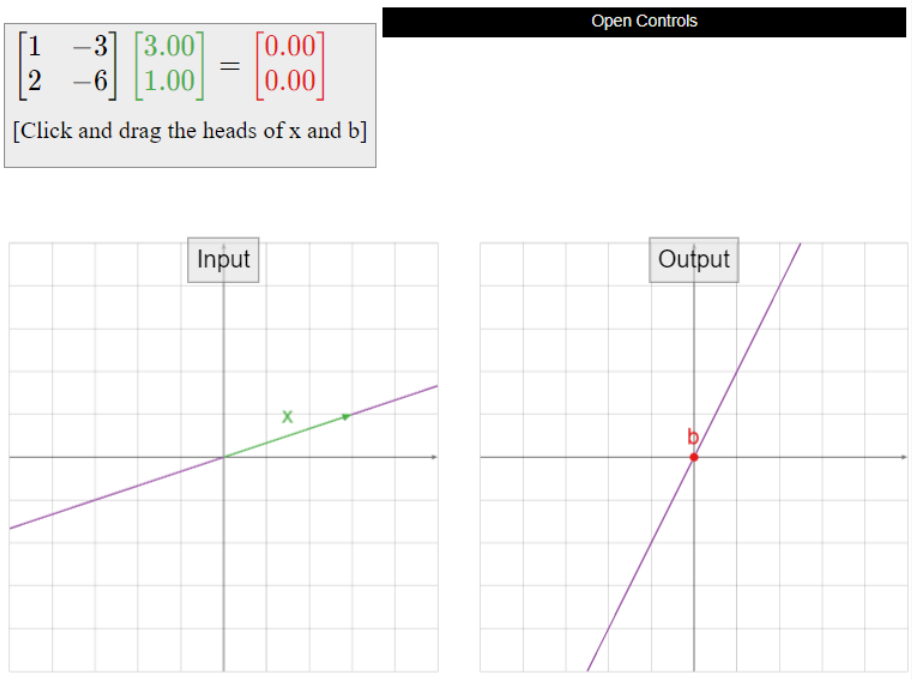


Figure 1.7.2: Interactive picture of the solution set of $Ax = 0$. If you drag x along the line spanned by $\begin{pmatrix}3\\1\end{pmatrix}$, the product Ax is always equal to zero. This is what it means for $\text{Span}\{\begin{pmatrix}3\\1\end{pmatrix}\}$ to be the solution set of $Ax = 0$.

Since there were *two* variables in the above Example 1.7.3, the solution set is a subset of $\mathbb{R}^2$. Since *one* of the variables was free, the solution set is a *line*:

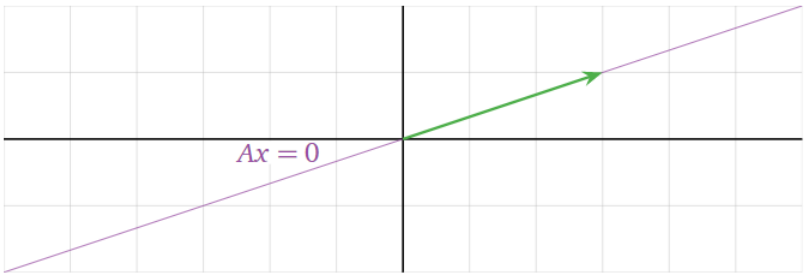


Figure 1.7.3

Note 1.7.3

In order to actually *find* a nontrivial solution to $Ax = 0$ in the above Example 1.7.3, it suffices to substitute any nonzero value for the free variable x_2. For instance, taking $x_2 = 1$ gives the nontrivial solution $x = 1 \cdot \begin{pmatrix}3\\1\end{pmatrix} = \begin{pmatrix}3\\1\end{pmatrix}$. Compare to this note in Section 1.3, Note 1.3.1.

Example 1.7.4: The solution set is a plane

Compute the parametric vector form of the solution set of $Ax = 0$, where

$$A = \begin{pmatrix}1 & -1 & 2\\-2 & 2 & -4\end{pmatrix}.$$

Solution

We row reduce (without augmenting, as suggested in the above Observation 1.7.2):

$$\begin{pmatrix} 1 & -1 & 2 \\ -2 & 2 & -4 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & -1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

This corresponds to the single equation $x_1 - x_2 + 2x_3 = 0$. We write the parametric form including the redundant equations $x_2 = x_2$ and $x_3 = x_3$:

$$\begin{cases} x_1 &= x_2 &- 2x_3 \\ x_2 &= x_2 & \\ x_3 &= & x_3. \end{cases}$$

We turn these into a single vector equation:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}.$$

This is the parametric vector form of the solution set. Since x_2 and x_3 are allowed to be anything, this says that the solution set is the set of all linear combinations of $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$. In other words, the solution set is

$$\text{Span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

We know how to draw the span of two noncollinear vectors in $\mathbb{R}^3$: it is a plane. Therefore, this is a picture of the solution set:

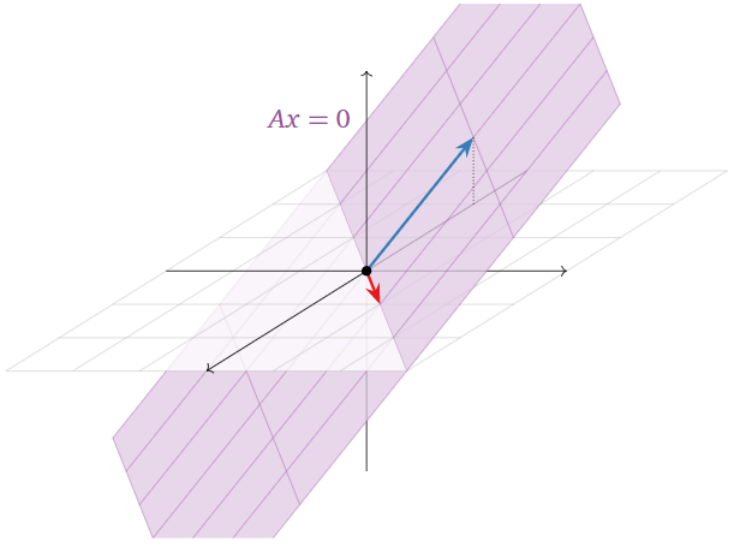


Figure 1.7.4

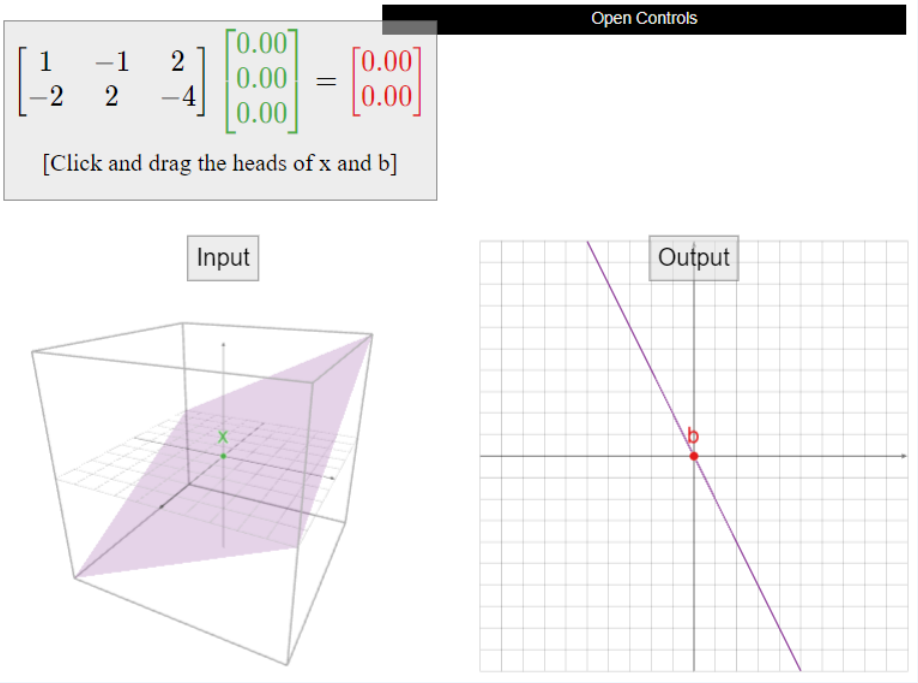


Figure 1.7.5: Interactive picture of the solution set of $Ax = 0$. If you drag x along the violet plane, the product Ax is always equal to zero. This is what it means for the plane to be the solution set of $Ax = 0$.

Since there were *three* variables in the above Example 1.7.4, the solution set is a subset of $\mathbb{R}^3$. Since *two* of the variables were free, the solution set is a *plane*.

There is a natural question to ask here: is it possible to write the solution to a homogeneous matrix equation using fewer vectors than the one given in the above recipe? We will see in Example 2.5.3 in Section 2.5 that the answer is *no*: the vectors from the recipe are always linearly independent, which means that there is no way to write the solution with fewer vectors.

Another natural question is: are the solution sets for inhomogeneuous equations also spans? As we will see shortly, they are never spans, but they are closely related to spans.

There is a natural relationship between the number of free variables and the "size" of the solution set, as follows.

Note 1.7.4: Dimension of the Solution Set

The above examples show us the following pattern: when there is one free variable in a consistent matrix equation, the solution set is a line, and when there are two free variables, the solution set is a plane, etc. The number of free variables is called the *dimension* of the solution set.

We will develop a rigorous definition, Definition 2.7.2, of dimension in Section 2.7, but for now the dimension will simply mean the number of free variables. Compare with this important note in Section 2.5, Note 2.5.5.

Intuitively, the dimension of a solution set is the number of parameters you need to describe a point in the solution set. For a line only one parameter is needed, and for a plane two parameters are needed. This is similar to how the location of a building on Peachtree Street—which is like a line—is determined by one number and how a street corner in Manhattan—which is like a plane—is specified by two numbers.

Inhomogeneous Systems

Recall that a matrix equation $Ax = b$ is called ***inhomogeneous*** when $b \neq 0$.

Example 1.7.5: The solution set is a line

What is the solution set of $Ax = b$, where

$$A = \begin{pmatrix} 1 & -3 \\ 2 & -6 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} -3 \\ -6 \end{pmatrix}?$$

(Compare to this Example 1.7.4, where we solved the corresponding homogeneous equation.)

Solution

We row reduce the associated augmented matrix:

$$\left(\begin{array}{cc|c} 1 & -3 & -3 \\ 2 & -6 & -6 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & -3 & -3 \\ 0 & 0 & 0 \end{array}\right).$$

This corresponds to the single equation $x_1 - 3x_2 = -3$. We can write the parametric form as follows:

$$\begin{cases} x_1 &= 3x_2 &- 3 \\ x_2 &= x_2 &+ 0. \end{cases}$$

We turn the above system into a *vector equation*:

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_2 \begin{pmatrix} 3 \\ 1 \end{pmatrix} + \begin{pmatrix} -3 \\ 0 \end{pmatrix}.$$

This vector equation is called the ***parametric vector form*** of the solution set. We write the solution set as

$$\text{Span}\left\{\begin{pmatrix} 3 \\ 1 \end{pmatrix}\right\} + \begin{pmatrix} -3 \\ 0 \end{pmatrix}.$$

Here is a picture of the the solution set:

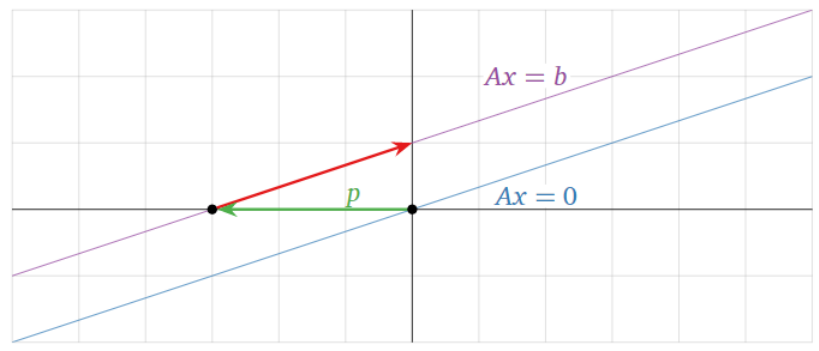


Figure 1.7.6

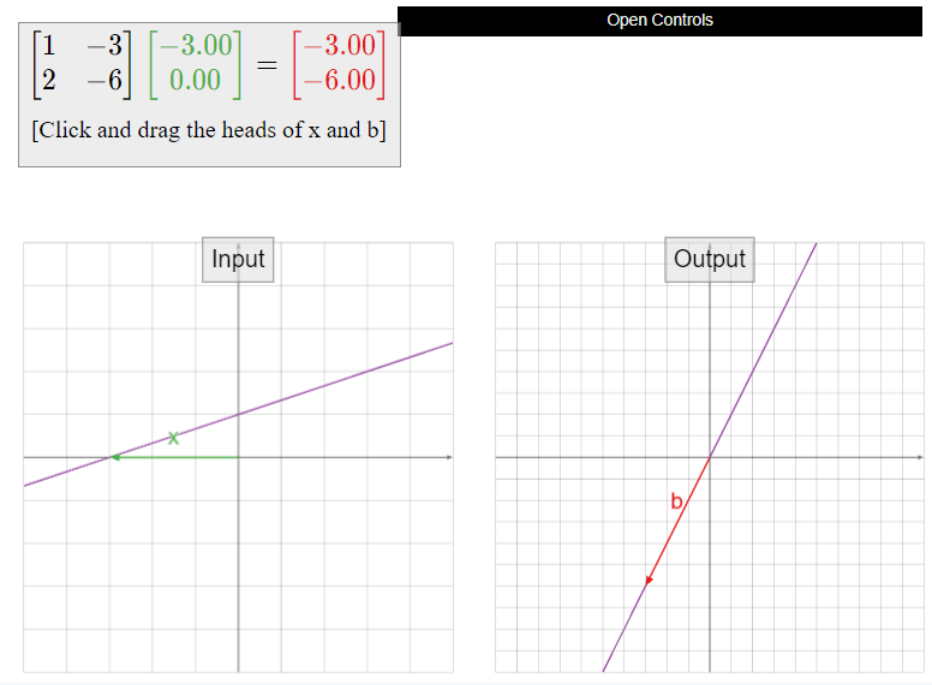


Figure 1.7.7: Interactive picture of the solution set of $Ax = b$. If you drag x along the violet line, the product Ax is always equal to b. This is what it means for the line to be the solution set of $Ax = b$.

In the above Example 1.7.5, the solution set was all vectors of the form

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_2 \begin{pmatrix} 3 \\ 1 \end{pmatrix} + \begin{pmatrix} -3 \\ 0 \end{pmatrix}$$

where x_2 is any scalar. The vector $p = \begin{pmatrix} -3 \\ 0 \end{pmatrix}$ is also a solution of $Ax = b$: take $x_2 = 0$. We call p a ***particular solution***.

In the solution set, x_2 is allowed to be anything, and so the solution set is obtained as follows: we take all scalar multiples of $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ and then add the particular solution $p = \begin{pmatrix} -3 \\ 0 \end{pmatrix}$ to each of these scalar multiples. Geometrically, this is accomplished by first drawing the span of $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$, which is a line through the origin (and, not coincidentally, the solution to $Ax = 0$), and we *translate*, or push, this line along $p = \begin{pmatrix} -3 \\ 0 \end{pmatrix}$. The translated line contains p and is parallel to $\text{Span}\{\begin{pmatrix} 3 \\ 1 \end{pmatrix}\}$: it is a *translate of a line*.

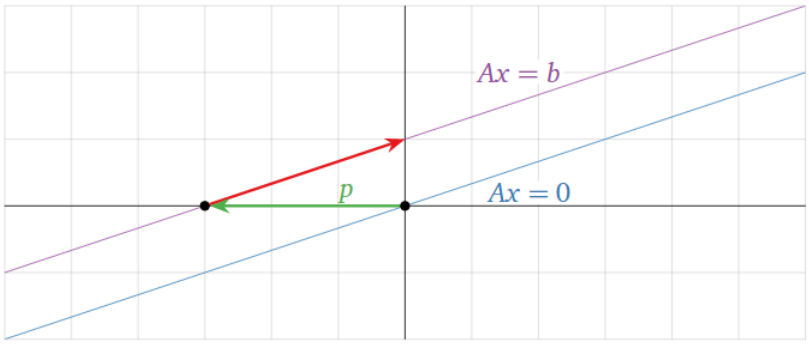


Figure 1.7.8

Example 1.7.6: The solution set is a plane

What is the solution set of $Ax = b$, where

$$A = \begin{pmatrix} 1 & -1 & 2 \\ -2 & 2 & -4 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 1 \\ -2 \end{pmatrix}?$$

(Compare this Example 1.7.4.)

Solution

We row reduce the associated augmented matrix:

$$\left(\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ -2 & 2 & -4 & -2 \end{array}\right) \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & -1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

This corresponds to the single equation $x_1 - x_2 + 2x_3 = 1$. We can write the parametric form as follows:

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

We turn the above system into a *vector equation*:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

This vector equation is called the ***parametric vector form*** of the solution set. Since x_2 and x_3 are allowed to be anything, this says that the solution set is the set of all linear combinations of $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$, *translated* by the vector $p = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. This is a plane which contains p and is parallel to $\text{Span}\left\{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}\right\}$: it is a *translate of a plane*. We write the solution set as

$$\text{Span}\left\{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}\right\} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Here is a picture of the solution set:

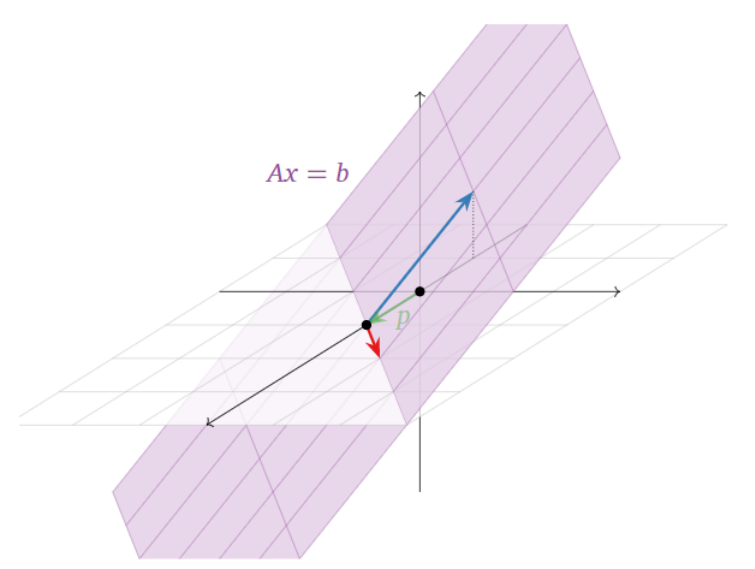


Figure 1.7.9

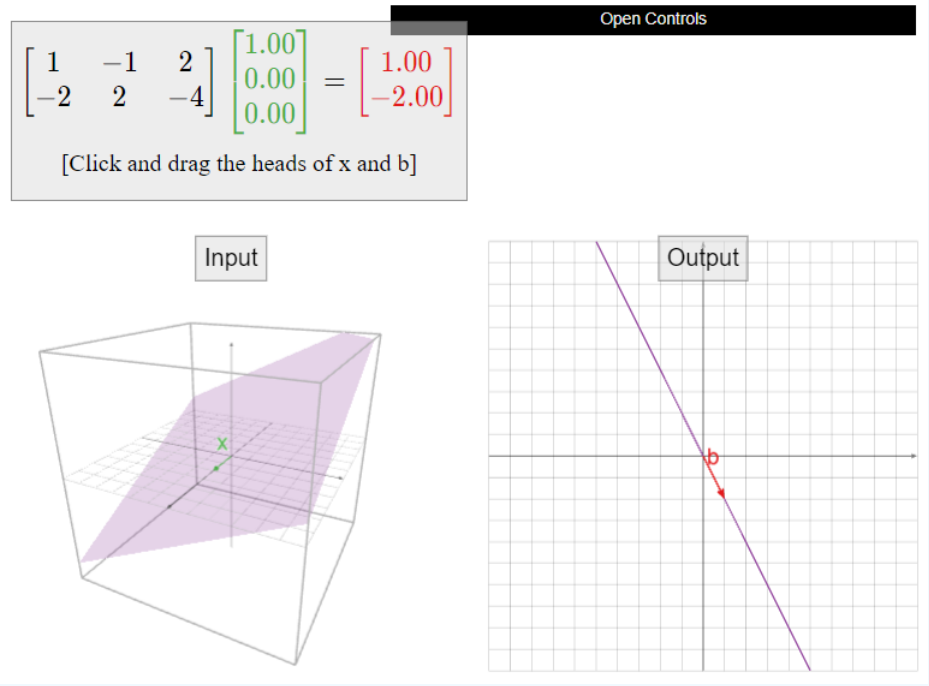


Figure 1.7.10: Interactive picture of the solution set of $Ax = b$. If you drag x along the violet plane, the product Ax is always equal to b. This is what it means for the plane to be the solution set of $Ax = b$.

In the above Example 1.7.6, the solution set was all vectors of the form

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

where x_2 and x_3 are any scalars. In this case, a particular solution is $p = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

In the previous Example 1.7.6 and the Example 1.7.5 before it, the parametric vector form of the solution set of $Ax = b$ was exactly the same as the parametric vector form of the solution set of $Ax = 0$ (from this Example 1.7.4 and this Example 1.7.3, respectively), plus a particular solution.

Key Observation 1.7.3

If $Ax = b$ is consistent, the set of solutions to is obtained by taking one ***particular solution*** p of $Ax = b$, and adding all solutions of $Ax = 0$.

In particular, if $Ax = b$ is consistent, the solution set is a ***translate of a span***.

The ***parametric vector form*** of the solutions of $Ax = b$ is just the parametric vector form of the solutions of $Ax = 0$, plus a particular solution p.

It is not hard to see why the key observation 1.7.3 is true. If p is a particular solution, then $Ap = b$, and if x is a solution to the homogeneous equation $Ax = 0$, then

$$A(x + p) = Ax + Ap = 0 + b = b,$$

so $x + p$ is another solution of $Ax = b$. On the other hand, if we start with any solution x to $Ax = b$ then $x - p$ is a solution to $Ax = 0$ since

$$A(x - p) = Ax - Ap = b - b = 0.$$

Remark

Row reducing to find the parametric vector form will give you one particular solution p of $Ax = b$. But the key observation 1.7.3 is true for any solution p. In other words, if we row reduce in a different way and find a different solution p' to $Ax = b$ then the solutions to $Ax = b$ can be obtained from the solutions to $Ax = 0$ by either adding p or by adding p'.

Example 1.7.7: The solution set is a point

What is the solution set of $Ax = b$, where

$$A = \begin{pmatrix} 1 & 3 & 4 \\ 2 & -1 & 2 \\ 1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}?$$

Solution

We form an augmented matrix and row reduce:

$$\left(\begin{array}{ccc|c} 1 & 3 & 4 & 0 \\ 2 & -1 & 2 & 1 \\ 1 & 0 & 1 & 0 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{array}\right).$$

The only solution is $p = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$.

According to the key observation 1.7.3, this is supposed to be a translate of a span by p. Indeed, we saw in the first Example 1.7.1 that the only solution of $Ax = 0$ is the trivial solution, i.e., that the solution set is the one-point set $\{0\}$. The solution set of the inhomogeneous equation $Ax = b$ is

$$\{0\} + \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}.$$

Note that $\{0\} = \text{Span}\{0\}$, so the homogeneous solution set is a span.

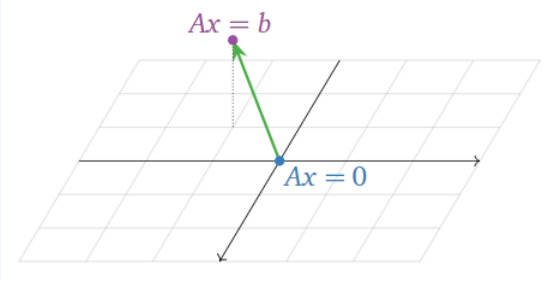


Figure 1.7.11

See the interactive figures in the next Subsection Solution Sets and Column Spans for visualizations of the key observation 1.7.3.

Note 1.7.5: Dimension of the Solution Set

As in this important note, when there is one free variable in a consistent matrix equation, the solution set is a line—this line does not pass through the origin when the system is inhomogeneous—when there are two free variables, the solution set is a plane (again not through the origin when the system is inhomogeneous), etc.

Again compare with this important note in Section 2.5, Note 2.5.5.

Solution Sets and Column Spans

To every $m \times n$ matrix A, we have now associated two completely different geometric objects, both described using spans.

- The ***solution set***: for fixed b, this is the set of all x such that $Ax = b$.
  - This is a span if $b = 0$, and it is a translate of a span if $b \neq 0$ (and $Ax = b$ is consistent).
  - It is a subset of $\mathbb{R}^n$.
  - It is computed by solving a system of equations: usually by row reducing and finding the parametric vector form.
- The ***span of the columns*** *of* A: this is the set of all b such that $Ax = b$ is consistent.
  - This is always a span.
  - It is a subset of $\mathbb{R}^m$.
  - It is not computed by solving a system of equations: row reduction plays no role.

Note 1.7.6

Do not confuse these two geometric constructions! In the first the question is which x's work for a given b and in the second the question is which b's work for some x.

Example 1.7.8: Interactive: Solution set and span of the columns (1)

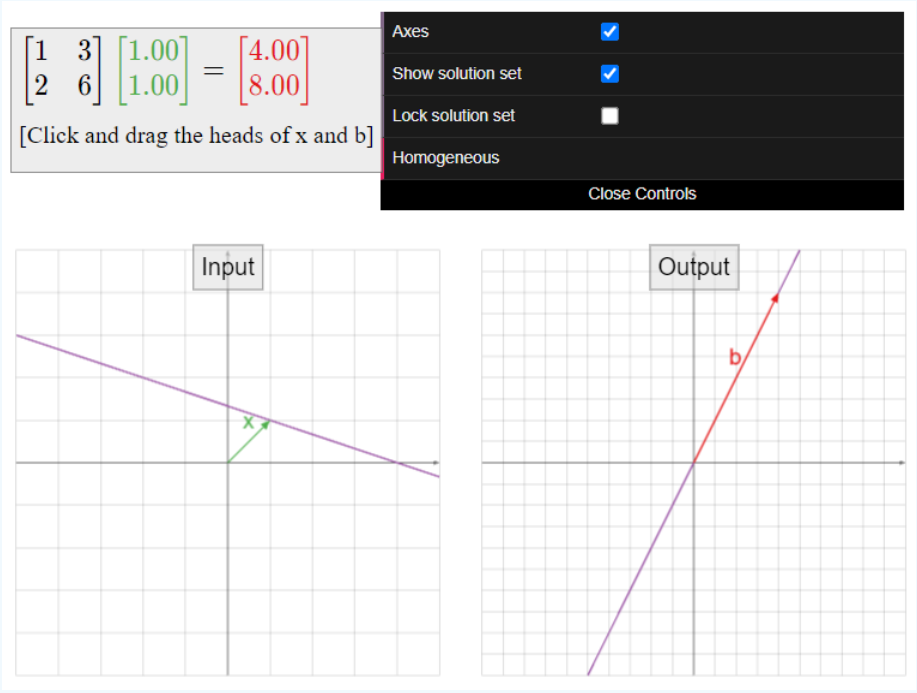


Figure 1.7.12: Left: the solution set of $Ax = b$ is in violet. Right: the span of the columns of A is in violet. As you move x, you change b, so the solution set changes—but all solution sets are parallel planes. If you move b within the span of the columns, the solution set also changes, and the demo solves the equation to find a particular solution x. If you move b outside of the span of the columns, the system becomes inconsistent, and the solution set disappears.

Example 1.7.9: Interactive: Solution set and span of the columns (2)

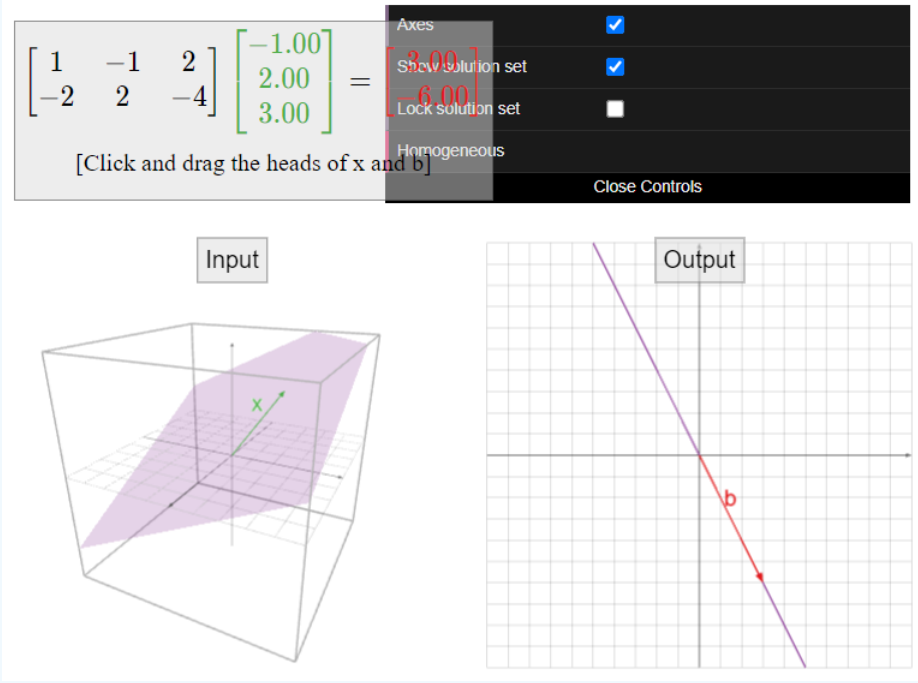


Figure 1.7.13: Left: the solution set of $Ax = b$ is in violet. Right: the span of the columns of A is in violet. As you move x, you change b, so the solution set changes—but all solution sets are parallel planes. If you move b within the span of the columns, the solution set also changes, and the demo solves the equation to find a particular solution x. If you move b outside of the span of the columns, the system becomes inconsistent, and the solution set disappears.

Example 1.7.10: Interactive: Solution set and span of the columns (3)

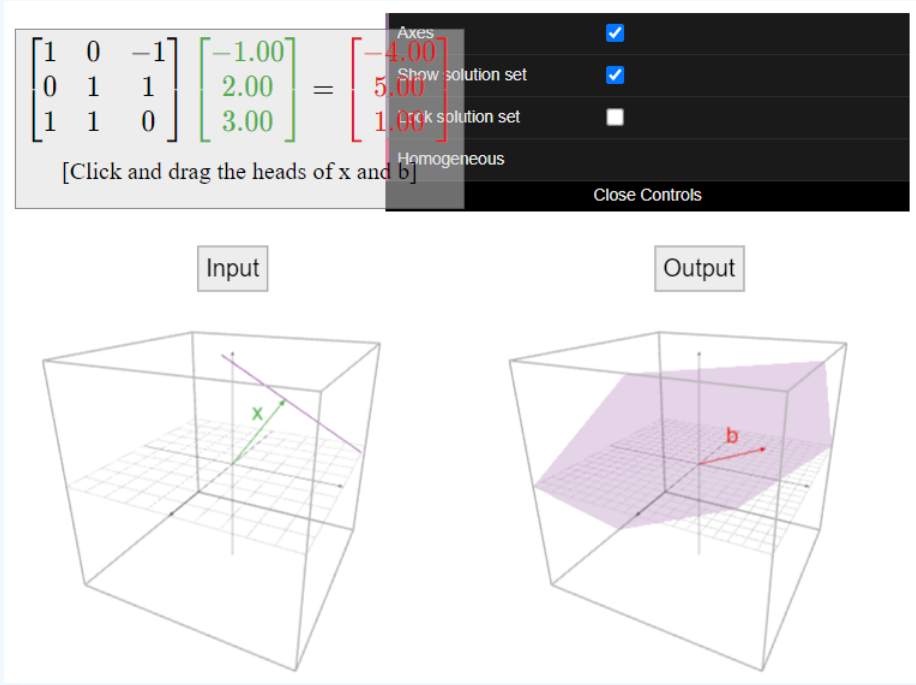


Figure 1.7.14: Left: the solution set of $Ax = b$ is in violet. Right: the span of the columns of A is in violet. As you move x, you change b, so the solution set changes—but all solution sets are parallel planes. If you move b within the span of the columns, the solution set also changes, and the demo solves the equation to find a particular solution x. If you move b outside of the span of the columns, the system becomes inconsistent, and the solution set disappears.

This page titled 1.7: Solution Sets is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **2.4: Solution Sets** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

1.8: Applications of Linear Systems

(I) An Application to Resistor Networks

The tools of linear algebra can be used to study the application of resistor networks. An example of an electrical circuit is below.

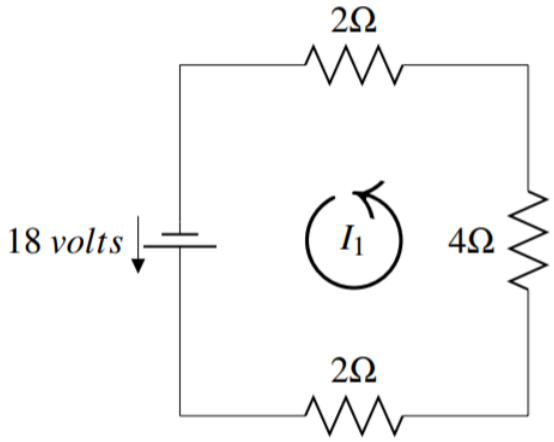


Figure 1.8.1

The jagged lines (—/\/\/\—) denote resistors and the numbers next to them give their resistance in ohms, written as Ω. The voltage source (—|| —) causes the current to flow in the direction from the shorter of the two lines toward the longer (as indicated by the arrow). The current for a circuit is labeled I_k.

In the above figure, the current I_1 has been labeled with an arrow in the counter clockwise direction. This is an entirely arbitrary decision and we could have chosen to label the current in the clockwise direction. With our choice of direction here, we define a positive current to flow in the counter clockwise direction and a negative current to flow in the clockwise direction.

The goal of this section is to use the values of resistors and voltage sources in a circuit to determine the current. An essential theorem for this application is Kirchhoff's law.

Theorem 1.8.1: Kirchhoff's Law

The sum of the resistance (R) times the amps (I) in the counter clockwise direction around a loop equals the sum of the voltage sources (V) in the same direction around the loop.

Kirchhoff's law allows us to set up a system of linear equations and solve for any unknown variables. When setting up this system, it is important to trace the circuit in the counter clockwise direction. If a resistor or voltage source is crossed against this direction, the related term must be given a negative sign.

We will explore this in the next example where we determine the value of the current in the initial diagram.

Example 1.8.1: Solving for Current

Applying Kirchhoff's Law to the diagram below, determine the value for I_1.

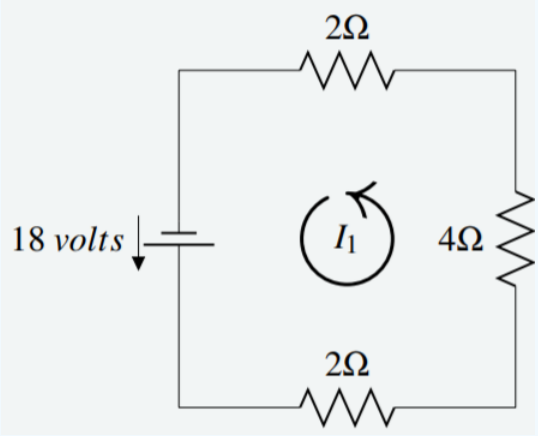


Figure 1.8.2

Solution

Begin in the bottom left corner, and trace the circuit in the counter clockwise direction. At the first resistor, multiplying resistance and current gives $2I_1$. Continuing in this way through all three resistors gives $2I_1+4I_1+2I_1$. This must equal the voltage source in the same direction. Notice that the direction of the voltage source matches the counter clockwise direction specified, so the voltage is positive.

Therefore the equation and solution are given by

$$\begin{aligned} 2I_1+4I_1+2I_1 &= 18 \\ 8I_1 &= 18 \\ I_1 &= \frac{9}{4}A \end{aligned}$$

Since the answer is positive, this confirms that the current flows counter clockwise.

Example 1.8.2 Solving for Current

Applying Kirchhoff's Law to the diagram below, determine the value for I_1.

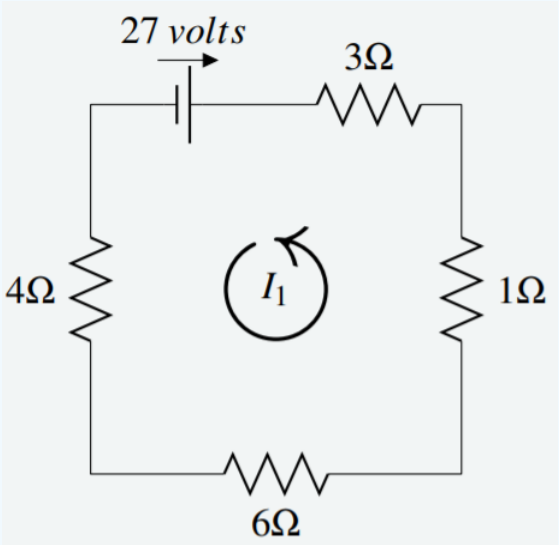


Figure 1.8.3

Solution

Begin in the top left corner this time, and trace the circuit in the counter clockwise direction. At the first resistor, multiplying resistance and current gives $4I_1$. Continuing in this way through the four resistors gives $4I_1+6I_1+1I_1+3I_1$. This must

equal the voltage source in the same direction. Notice that the direction of the voltage source is opposite to the counter clockwise direction, so the voltage is negative.

Therefore the equation and solution are given by

$$\begin{aligned} 4I_1+6I_1+1I_1+3I_1 &= -27 \\ 14I_1 &= -27 \\ I_1 &= -\frac{27}{14}A \end{aligned}$$

Since the answer is negative, this tells us that the current flows clockwise.

A more complicated example follows. Two of the circuits below may be familiar; they were examined in the examples above. However as they are now part of a larger system of circuits, the answers will differ.

Example 1.8.3: Unknown Currents

The diagram below consists of four circuits. The current (I_k) in the four circuits is denoted by I_1, I_2, I_3, I_4. Using Kirchhoff's Law, write an equation for each circuit and solve for each current.

Solution

The circuits are given in the following diagram.

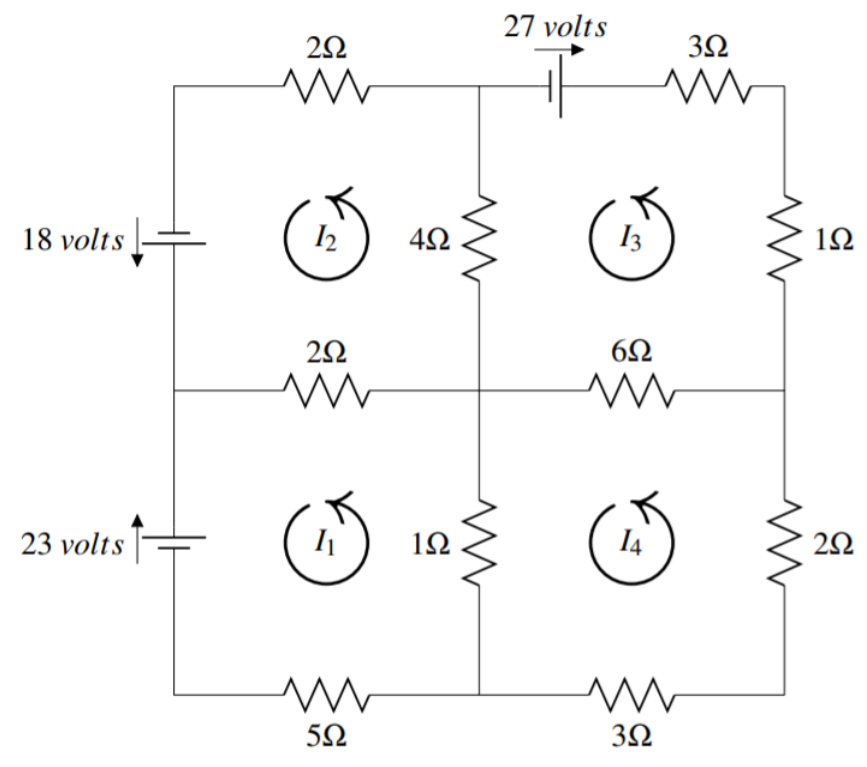


Figure 1.8.4

Starting with the top left circuit, multiply the resistance by the amps and sum the resulting products. Specifically, consider the resistor labeled 2Ω that is part of the circuits of I_1 and I_2. Notice that current I_2 runs through this in a positive (counter clockwise) direction, and I_1 runs through in the opposite (negative) direction. The product of resistance and amps is then $2(I_2-I_1)=2I_2-2I_1$. Continue in this way for each resistor, and set the sum of the products equal to the voltage source to write the equation:

$$2I_2-2I_1+4I_2-4I_3+2I_2=18$$

The above process is used on each of the other three circuits, and the resulting equations are:

Upper right circuit:

$$4I_3 - 4I_2 + 6I_3 - 6I_4 + I_3 + 3I_3 = -27$$

Lower right circuit:

$$3I_4 + 2I_4 + 6I_4 - 6I_3 + I_4 - I_1 = 0$$

Lower left circuit:

$$5I_1 + I_1 - I_4 + 2I_1 - 2I_2 = -23$$

Notice that the voltage for the upper right and lower left circuits are negative due to the clockwise direction they indicate.

The resulting system of four equations in four unknowns is

$$\begin{aligned} 2I_2 - 2I_1 + 4I_2 - 4I_3 + 2I_2 &= 18 \\ 4I_3 - 4I_2 + 6I_3 - 6I_4 + I_3 + I_3 &= -27 \\ 2I_4 + 3I_4 + 6I_4 - 6I_3 + I_4 - I_1 &= 0 \\ 5I_1 + I_1 - I_4 + 2I_1 - 2I_2 &= -23 \end{aligned}$$

Simplifying and rearranging with variables in order, we have:

$$\begin{aligned} -2I_1 + 8I_2 - 4I_3 &= 18 \\ -4I_2 + 14I_3 - 6I_4 &= -27 \\ -I_1 - 6I_3 + 12I_4 &= 0 \\ 8I_1 - 2I_2 - I_4 &= -23 \end{aligned}$$

The augmented matrix is

$$\left[\begin{array}{rrrr|r} -2 & 8 & -4 & 0 & 18 \\ 0 & -4 & 14 & -6 & -27 \\ -1 & 0 & -6 & 12 & 0 \\ 8 & -2 & 0 & -1 & -23 \end{array}\right]$$

The solution to this matrix is

$$\begin{aligned} I_1 &= -3A \\ I_2 &= \frac{1}{4}A \\ I_3 &= -\frac{5}{2}A \\ I_4 &= -\frac{3}{2}A \end{aligned}$$

This tells us that currents I_1, I_3, and I_4 travel clockwise while I_2 travels counter clockwise.

(II) Balancing Chemical Reactions

The tools of linear algebra can also be used in the subject area of Chemistry, specifically for balancing chemical reactions.

Consider the chemical reaction

$$SnO_2 + H_2 \rightarrow Sn + H_2O$$

Here the elements involved are tin (Sn), oxygen (O), and hydrogen (H). A chemical reaction occurs and the result is a combination of tin (Sn) and water (H_2O). When considering chemical reactions, we want to investigate how much of each element we began with and how much of each element is involved in the result.

An important theory we will use here is the mass balance theory. It tells us that we cannot create or delete elements within a chemical reaction. For example, in the above expression, we must have the same number of oxygen, tin, and hydrogen on both sides of the reaction. Notice that this is not currently the case. For example, there are two oxygen atoms on the left and only one on the right. In order to fix this, we want to find numbers x, y, z, w such that

$$xSnO_2 + yH_2 \rightarrow zSn + wH_2O$$

where both sides of the reaction have the same number of atoms of the various elements.

This is a familiar problem. We can solve it by setting up a system of equations in the variables x, y, z, w. Thus you need

$$\begin{aligned} Sn: &\quad x = z \\ O: &\quad 2x = w \\ H: &\quad 2y = 2w \end{aligned}$$

We can rewrite these equations as

$$\begin{aligned} Sn: &\quad x - z = 0 \\ O: &\quad 2x - w = 0 \\ H: &\quad 2y - 2w = 0 \end{aligned}$$

The augmented matrix for this system of equations is given by

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

The reduced row-echelon form of this matrix is

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

The solution is given by

$$\begin{aligned} x - \tfrac{1}{2}w = 0 \\ y - w = 0 \\ z - \tfrac{1}{2}w = 0 \end{aligned}$$

which we can write as

$$\begin{aligned} x = \tfrac{1}{2}t \\ y = t \\ z = \tfrac{1}{2}t \\ w = t \end{aligned}$$

For example, let $w = 2$ and this would yield $x = 1, y = 2$, and $z = 1$. We can put these values back into the expression for the reaction which yields

$$SnO_2 + 2H_2 \rightarrow Sn + 2H_2O$$

Observe that each side of the expression contains the same number of atoms of each element. This means that it preserves the total number of atoms, as required, and so the chemical reaction is balanced.

Consider another example.

Example 1.8.1: Balancing a Chemical Reaction

Potassium is denoted by K, oxygen by O, phosphorus by P and hydrogen by H. Consider the reaction given by

$$KOH + H_3PO_4 \rightarrow K_3PO_4 + H_2O$$

Balance this chemical reaction.

Solution

We will use the same procedure as above to solve this problem. We need to find values for x, y, z, w such that

$$xKOH + yH_3PO_4 \rightarrow zK_3PO_4 + wH_2O$$

preserves the total number of atoms of each element.

Finding these values can be done by finding the solution to the following system of equations.

$$\begin{aligned} K: &\quad x = 3z \\ O: &\quad x + 4y = 4z + w \\ H: &\quad x + 3y = 2w \\ P: &\quad y = z \end{aligned}$$

The augmented matrix for this system is

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

and the reduced row-echelon form is

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

The solution is given by

$$\begin{aligned} x - w &= 0 \\ y - \tfrac{1}{3}w &= 0 \\ z - \tfrac{1}{3}w &= 0 \end{aligned}$$

which can be written as

$$\begin{aligned} x &= t \\ y &= \tfrac{1}{3}t \\ z &= \tfrac{1}{3}t \\ w &= t \end{aligned}$$

Choose a value for t, say 3. Then $w = 3$ and this yields $x = 3, y = 1, z = 1$. It follows that the balanced reaction is given by

$$3KOH + 1H_3PO_4 \rightarrow 1K_3PO_4 + 3H_2O$$

Note that this results in the same number of atoms on both sides.

Of course these numbers you are finding would typically be the number of moles of the molecules on each side. Thus three moles of KOH added to one mole of H_3PO_4 yields one mole of K_3PO_4 and three moles of H_2O.

This page titled 1.8: Applications of Linear Systems is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.8: An Application to Resistor Networks** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.
- **1.6: Balancing Chemical Reactions** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

1.9: Linear Independence

Objectives

1. Understand the concept of linear independence.
2. Learn two criteria for linear independence.
3. Understand the relationship between linear independence and pivot columns / free variables.
4. *Recipe:* test if a set of vectors is linearly independent / find an equation of linear dependence.
5. *Picture:* whether a set of vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ is linearly independent or not.
6. *Vocabulary words:* ***linear dependence relation*** / ***equation of linear dependence***.
7. *Essential vocabulary words:* ***linearly independent***, ***linearly dependent***.

Sometimes the span of a set of vectors is "smaller" than you expect from the number of vectors, as in the picture below. This means that (at least) one of the vectors is redundant: it can be removed without affecting the span. In the present section, we formalize this idea in the notion of *linear independence*.

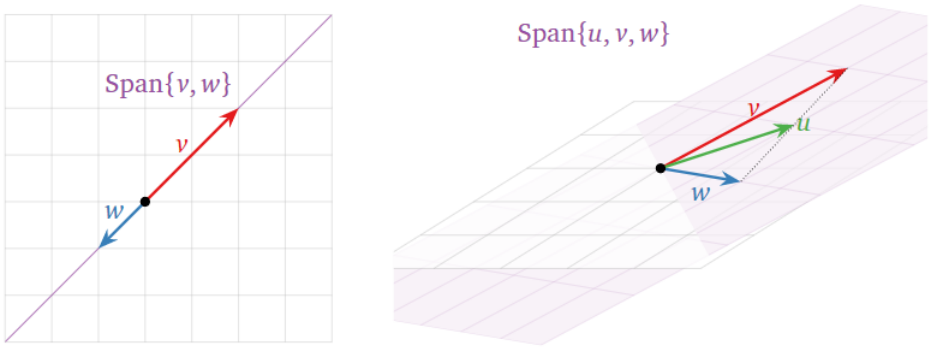


Figure 1.9.1: Pictures of sets of vectors that are linearly dependent. Note that in each case, one vector is in the span of the others—so it doesn't make the span bigger.

The Definition of Linear Independence

Definition 1.9.1: Linearly Independent and Linearly Dependent

A set of vectors $\{v_1, v_2, \ldots, v_k\}$ is ***linearly independent*** if the vector equation

$$x_1v_1 + x_2v_2 + \cdots + x_kv_k = 0$$

has only the trivial solution $x_1 = x_2 = \cdots = x_k = 0$. The set $\{v_1, v_2, \ldots, v_k\}$ is ***linearly dependent*** otherwise.

In other words, $\{v_1, v_2, \ldots, v_k\}$ is linearly dependent if there exist numbers $x_1, x_2, \ldots, x_k$, not all equal to zero, such that

$$x_1v_1 + x_2v_2 + \cdots + x_kv_k = 0.$$

This is called a ***linear dependence relation*** or ***equation of linear dependence***.

Note: 1.9.1

Note that linear dependence and linear independence are notions that apply to a *collection of vectors*. It does not make sense to say things like "this vector is linearly dependent on these other vectors," or "this matrix is linearly independent."

Example 1.9.1: Checking linear dependence

Is the set

$$\left\{\begin{pmatrix}1\\1\\1\end{pmatrix},\begin{pmatrix}1\\-1\\2\end{pmatrix},\begin{pmatrix}3\\1\\4\end{pmatrix}\right\}$$

linearly independent?

Solution

Equivalently, we are asking if the homogeneous vector equation

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

has a nontrivial solution. We solve this by forming a matrix and row reducing (we do not augment because of this Observation 2.4.2 in Section 2.4):

$$\begin{pmatrix}1&1&3\\1&-1&1\\1&2&4\end{pmatrix}\xrightarrow{\text{row reduce}}\begin{pmatrix}1&0&2\\0&1&1\\0&0&0\end{pmatrix}$$

This says $x=-2z$ and $y=-z$. So there exist nontrivial solutions: for instance, taking $z=1$ gives this equation of linear dependence:

$$-2\begin{pmatrix}1\\1\\1\end{pmatrix}-\begin{pmatrix}1\\-1\\2\end{pmatrix}+\begin{pmatrix}3\\1\\4\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}.$$

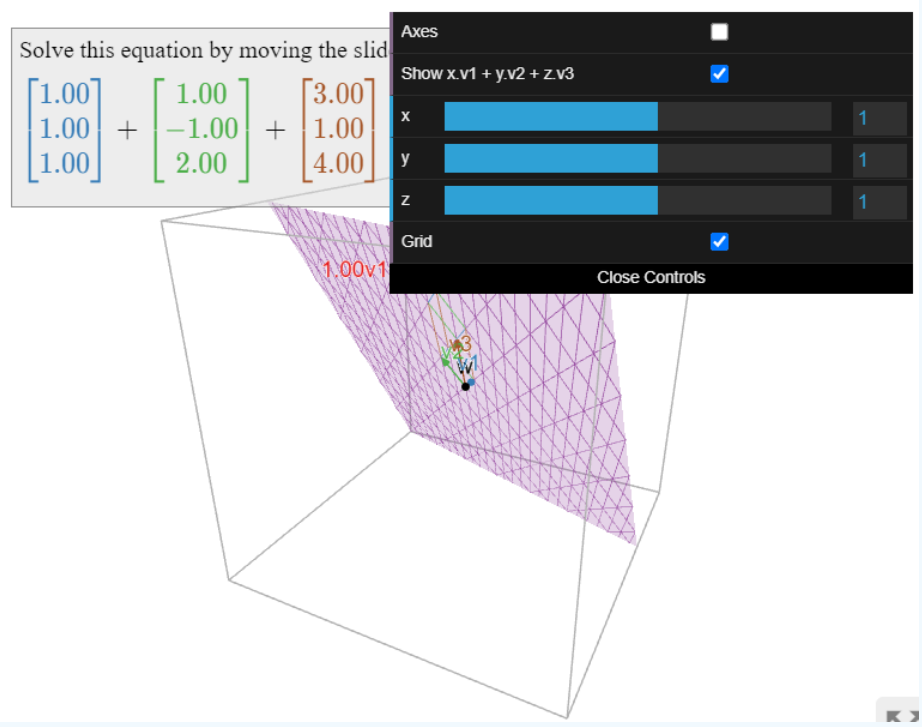


Figure 1.9.2: Move the sliders to solve the homogeneous vector equation in this example. Do you see why the vectors need to be coplanar in order for there to exist a nontrivial solution?

Example 1.9.2: Checking linear independence

Is the set

$$\left\{\begin{pmatrix}1\\1\\-2\end{pmatrix},\begin{pmatrix}1\\-1\\2\end{pmatrix},\begin{pmatrix}3\\1\\4\end{pmatrix}\right\}$$

linearly independent?

Solution

Equivalently, we are asking if the homogeneous vector equation

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

has a nontrivial solution. We solve this by forming a matrix and row reducing (we do not augment because of this Observation 2.4.2 in Section 2.4):

$$\begin{pmatrix}1&1&3\\1&-1&1\\-2&2&4\end{pmatrix}\xrightarrow{\text{row reduce}}\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$$

This says $x=y=z=0$, i.e., the only solution is the trivial solution. We conclude that the set is linearly independent.

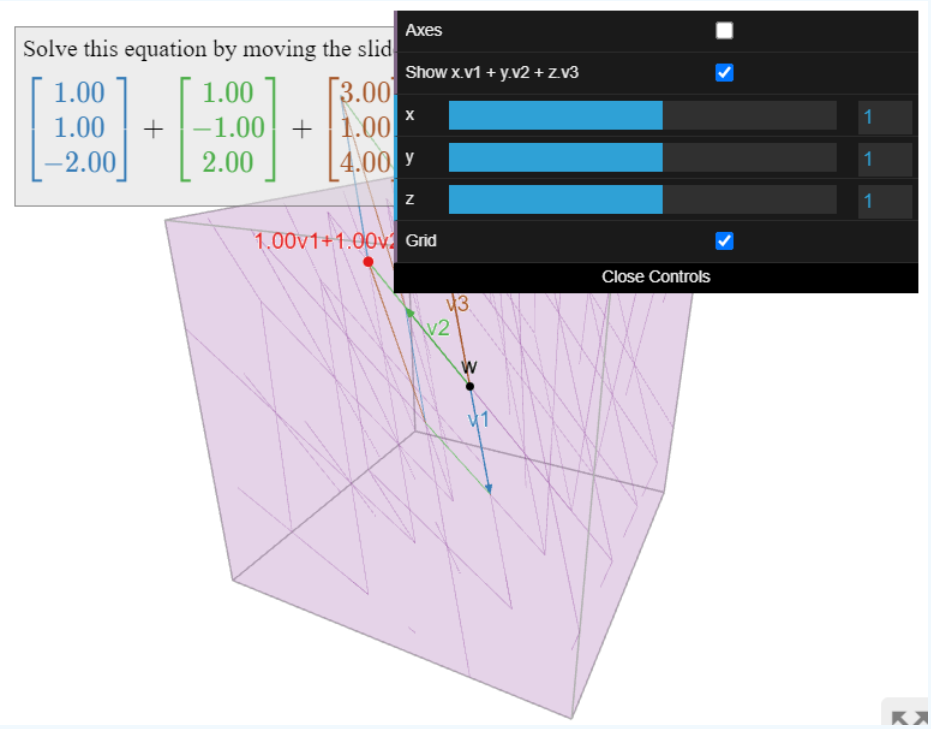


Figure 1.9.3: Move the sliders to solve the homogeneous vector equation in this example. Do you see why the vectors would need to be coplanar in order for there to exist a nontrivial solution?

Example 1.9.3: Vector parametric form

An important observation is that the vectors coming from the parametric vector form of the solution of a matrix equation $Ax=0$ are linearly independent. In Example 2.4.4 we saw that the solution set of $Ax=0$ for

$$A = \begin{pmatrix} 1 & -1 & 2 \\ -2 & 2 & -4 \end{pmatrix}?$$

is

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}.$$

Let's explain why the vectors $(1, 1, 0)$ and $(-2, 0, 1)$ are linearly independent. Suppose that

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = x_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x_2 - 2x_3 \\ x_2 \\ x_3 \end{pmatrix}.$$

Comparing the second and third coordinates, we see that $x_2 = x_3 = 0$. This reasoning will work in any example, since the entries corresponding to the free variables are all equal to 1 or 0, and are only equal to 1 in one of the vectors. This observation forms part of this theorem in Section 2.7, Theorem 2.7.2.

The above examples lead to the following recipe.

Recipe: Checking Linear Independence

A set of vectors $\{v_1, v_2, \ldots, v_k\}$ is linearly independent if and only if the vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_k v_k = 0$$

has only the trivial solution, if and only if the matrix equation $Ax = 0$ has only the trivial solution, where A is the matrix with columns $v_1, v_2, \ldots, v_k$:

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & & | \end{pmatrix}.$$

This is true if and only if A has a pivot position, Definition 1.2.5 in Section 1.2 in every column.

Solving the matrix equatiion $Ax = 0$ will either verify that the columns $v_1, v_2, \ldots, v_k$ are linearly independent, or will produce a linear dependence relation by substituting any nonzero values for the free variables.

(Recall that $Ax = 0$ has a nontrivial solution if and only if A has a column without a pivot: see this Observation 2.4.1 in Section 2.4.)

Suppose that A has more columns than rows. Then A cannot have a pivot in every column (it has at most one pivot per row), so its columns are automatically linearly dependent.

Note 1.9.2

A wide matrix (a matrix with more columns than rows) has linearly dependent columns.

For example, four vectors in $\mathbb{R}^3$ are automatically linearly dependent. Note that a tall matrix may or may not have linearly independent columns.

Fact 1.9.1: Facts About Linear Independence

1. Two vectors are linearly dependent if and only if they are collinear, i.e., one is a scalar multiple of the other.
2. Any set containing the zero vector is linearly dependent.
3. If a subset of $\{v_1, v_2, \ldots, v_k\}$ is linearly dependent, then $\{v_1, v_2, \ldots, v_k\}$ is linearly dependent as well.

Proof

1. If $v_1 = cv_2$ then $v_1 - cv_2 = 0$, so $\{v_1, v_2\}$ is linearly dependent. In the other direction, if $x_1v_1 + x_2v_2 = 0$ with $x_1 \neq 0$ (say), then $v_1 = -\frac{x_2}{x_1}v_2$.
2. It is easy to produce a linear dependence relation if one vector is the zero vector: for instance, if $v_1 = 0$ then
$$1 \cdot v_1 + 0 \cdot v_2 + \cdots + 0 \cdot v_k = 0.$$
3. After reordering, we may suppose that $\{v_1, v_2, \ldots, v_r\}$ is linearly dependent, with $r < p$. This means that there is an equation of linear dependence
$$x_1v_1 + x_2v_2 + \cdots + x_rv_r = 0,$$
with at least one of $x_1, x_2, \ldots, x_r$ nonzero. This is also an equation of linear dependence among $\{v_1, v_2, \ldots, v_k\}$,since we can take the coefficients of $v_{r+1}, \ldots, v_k$ to all be zero.

With regard to the first fact, note that the zero vector is a multiple of any vector, so it is collinear with any other vector. Hence facts 1 and 2 are consistent with each other.

Criteria for Linear Independence

In this subsection we give two criteria for a set of vectors to be linearly independent. Keep in mind, however, that the actual definition for linear independence, Definition 1.9.1, is above.

Theorem 1.9.1

A set of vectors $\{v_1, v_2, \ldots, v_k\}$ is linearly dependent if and only if one of the vectors is in the span of the other ones.

Any such vector may be removed without affecting the span.

Proof

Suppose, for instance, that v_3 is in $\text{Span}\{v_1, v_2, v_4\}$,so we have an equation like
$$v_3 = 2v_1 - \frac{1}{2}v_2 + 6v_4.$$
We can subract v_3 from both sides of the equation to get
$$0 = 2v_1 - \frac{1}{2}v_2 - v_3 + 6v_4.$$
This is a linear dependence relation.

In this case, any linear combination of v_1, v_2, v_3, v_4 is already a linear combination of v_1, v_2, v_4:
$$\begin{aligned} x_1v_1 + x_2v_2 + x_3v_3 + x_4v_4 &= x_1v_1 + x_2v_2 + x_3\left(2v_1 - \frac{1}{2}v_2 + 6v_4\right) + x_4v_4 \\ &= (x_1 + 2x_3)v_1 + \left(x_2 - \frac{1}{2}x_3\right)v_2 + (x_4 + 6)v_4. \end{aligned}$$
Therefore, $\text{Span}\{v_1, v_2, v_3, v_4\}$ is contained in $\text{Span}\{v_1, v_2, v_4\}$. Any linear combination of v_1, v_2, v_4 is also a linear combination of v_1, v_2, v_3, v_4 (with the v_3-coefficient equal to zero), so $\text{Span}\{v_1, v_2, v_4\}$ is also contained in $\text{Span}\{v_1, v_2, v_3, v_4\}$,and thus they are equal.

In the other direction, if we have a linear dependence relation like
$$0 = 2v_1 - \frac{1}{2}v_2 + v_3 - 6v_4,$$
then we can move any nonzero term to the left side of the equation and divide by its coefficient:
$$v_1 = \frac{1}{2}\left(\frac{1}{2}v_2 - v_3 + 6v_4\right).$$
This shows that v_1 is in $\text{Span}\{v_2, v_3, v_4\}$.

We leave it to the reader to generalize this proof for any set of vectors.

Note 1.9.3

In a linearly dependent set $\{v_1, v_2, \ldots, v_k\}$, it is not generally true that *any* vector v_j is in the span of the others, only that *at least one* of them is.

For example, the set $\left\{ \binom{1}{0}, \binom{2}{0}, \binom{0}{1} \right\}$ is linearly dependent, but $\binom{0}{1}$ is not in the span of the other two vectors. Also see this Figure 1.9.14below.

The previous Theorem 1.9.1 makes precise in what sense a set of linearly dependent vectors is redundant.

Theorem 1.9.2: Increasing Span Criterion

A set of vectors $\{v_1, v_2, \ldots, v_k\}$ is linearly independent if and only if, for every j, the vector v_j is not in $\text{Span}\{v_1, v_2, \ldots, v_{j-1}\}$.

Proof

It is equivalent to show that $\{v_1, v_2, \ldots, v_k\}$ is linearly dependent if and only if v_j is in $\text{Span}\{v_1, v_2, \ldots, v_{j-1}\}$ for some j. The "if" implication is an immediate consequence of the previous Theorem 1.9.1 Suppose then that $\{v_1, v_2, \ldots, v_k\}$ is linearly dependent. This means that some v_j is in the span of the others. Choose the largest such j. We claim that this v_j is in $\text{Span}\{v_1, v_2, \ldots, v_{j-1}\}$. If not, then

$$v_j = x_1 v_1 + x_2 v_2 + \cdots + x_{j-1} v_{j-1} + x_{j+1} v_{j+1} + \cdots + x_k v_k$$

with not all of $x_{j+1}, \ldots, x_k$ equal to zero. Suppose for simplicity that $x_k \neq 0$. Then we can rearrange:

$$v_k = -\frac{1}{x_k}\left(x_1 v_1 + x_2 v_2 + \cdots + x_{j-1} v_{j-1} - v_j + x_{j+1} v_{j+1} + \cdots + x_{p-1} v_{p-1}\right).$$

This says that v_k is in the span of $\{v_1, v_2, \ldots, v_{p-1}\}$, which contradicts our assumption that v_j is the last vector in the span of the others.

We can rephrase this as follows:

Note 1.9.4

If you make a set of vectors by adding one vector at a time, and if the span got bigger every time you added a vector, then your set is linearly independent.

Pictures of Linear Independence

A set containg one vector $\{v\}$ is linearly independent when $v \neq 0$, since $xv = 0$ implies $x = 0$.

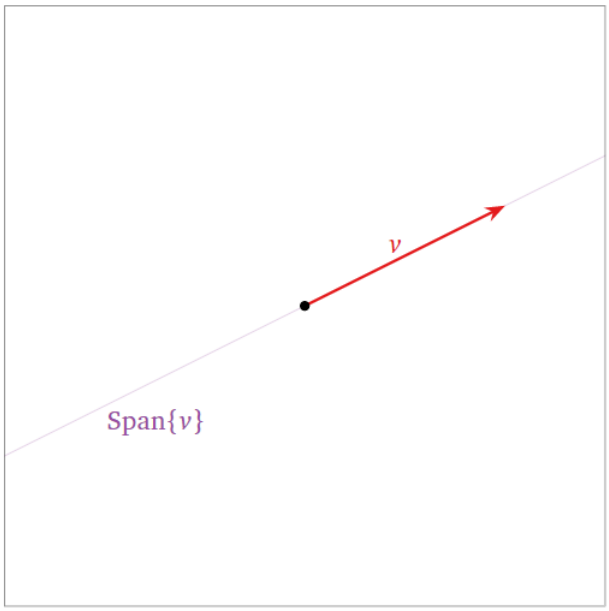


Figure 1.9.4

A set of two noncollinear vectors $\{v, w\}$ is linearly independent:

- Neither is in the span of the other, so we can apply the first criterion, Theorem 1.9.1.
- The span got bigger when we added w, so we can apply the increasing span criterion, Theorem 1.9.2.

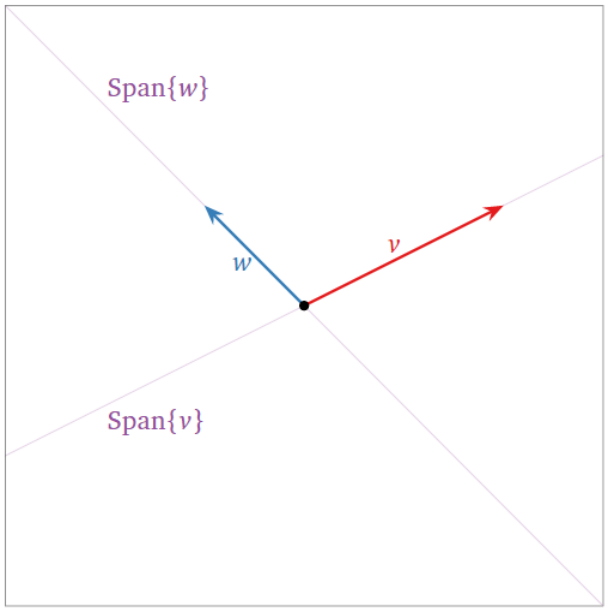


Figure 1.9.5

The set of three vectors $\{v, w, u\}$ below is linearly dependent:

- u is in $\text{Span}\{v, w\}$, so we can apply the first criterion, Theorem 1.9.1.
- The span did not increase when we added u, so we can apply the increasing span criterion, Theorem 1.9.2.

In the picture below, note that v is in $\text{Span}\{u, w\}$, and w is in $\text{Span}\{u, v\}$, so we can remove any of the three vectors without shrinking the span.

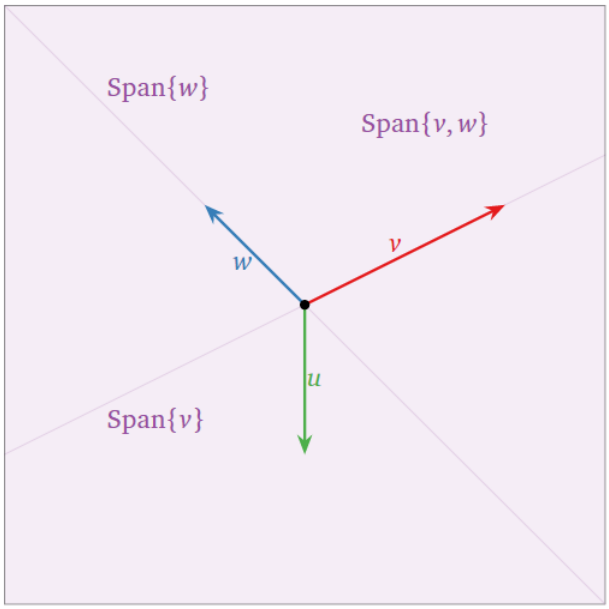


Figure 1.9.6

Two collinear vectors are always linearly dependent:

- w is in $\text{Span}\{v\}$, so we can apply the first criterion, Theorem 1.9.1.
- The span did not increase when we added w, so we can apply the increasing span criterion, Theorem 1.9.2.

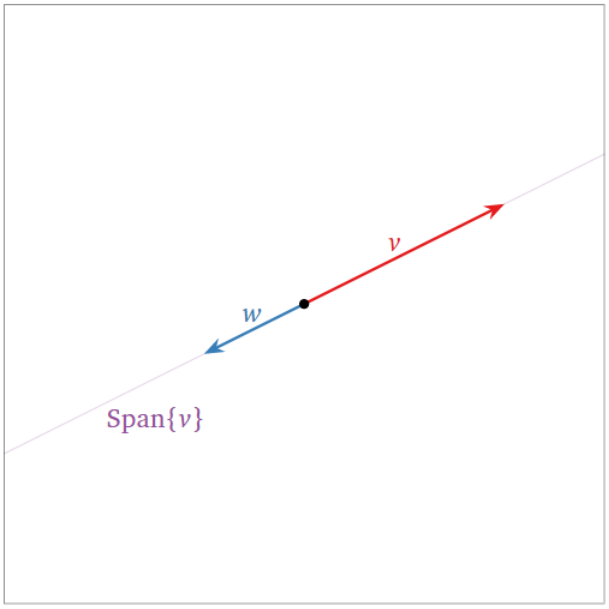


Figure 1.9.7

These three vectors $\{v, w, u\}$ are linearly dependent: indeed, $\{v, w\}$ is already linearly dependent, so we can use the third Fact 1.9.1.

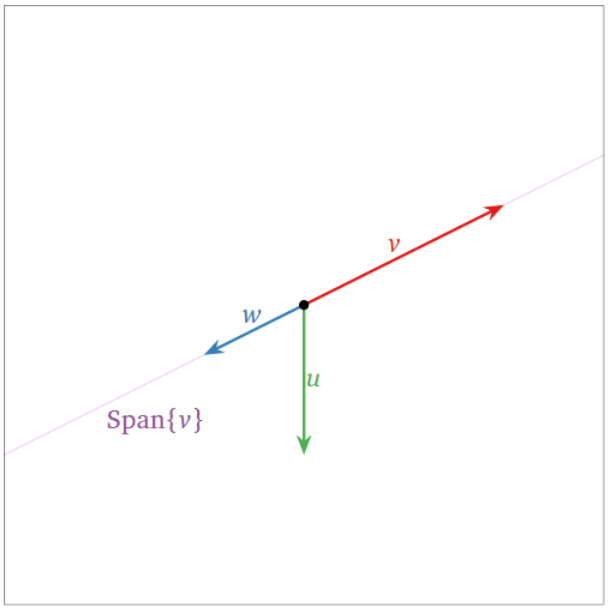


Figure 1.9.8

Example 1.9.4: Interactive: Linear independence of two vectors in $\mathbb{R}^2$

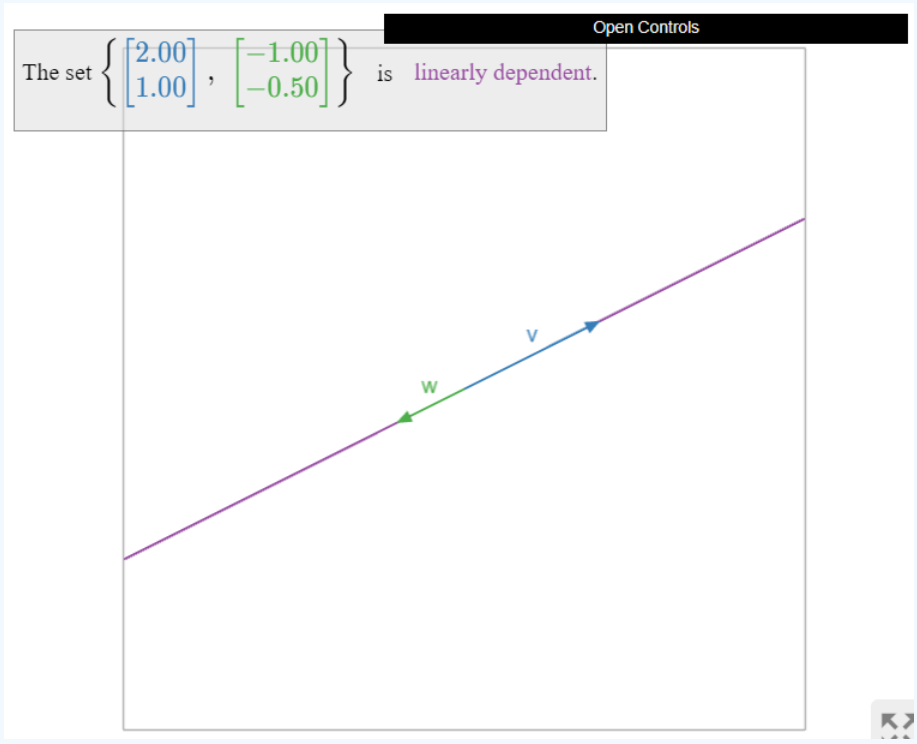


Figure 1.9.9: Move the vector heads and the demo will tell you if they are linearly independent and show you their span.

Example 1.9.5: Interactive: Linear dependence of three vectors in $\mathbb{R}^2$

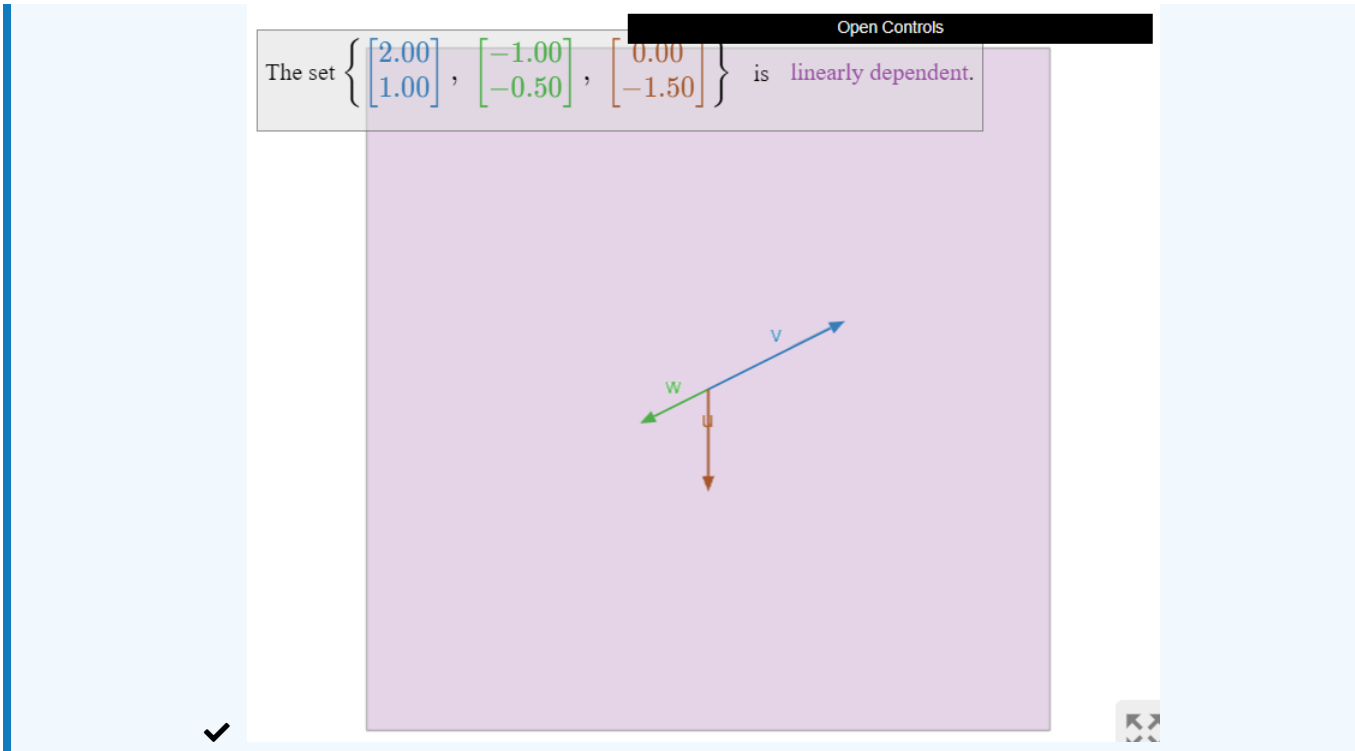


Figure 1.9.10: Move the vector heads and the demo will tell you that they are linearly dependent and show you their span.

The two vectors $\{v, w\}$ below are linearly independent because they are not collinear.

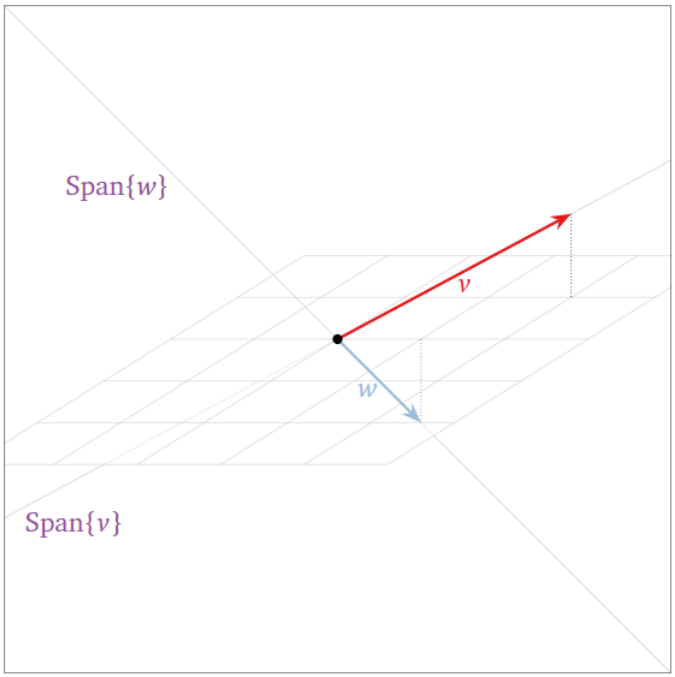


Figure 1.9.11

The three vectors $\{v, w, u\}$ below are linearly independent: the span got bigger when we added w, then again when we added u, so we can apply the increasing span criterion, Theorem 1.9.2.

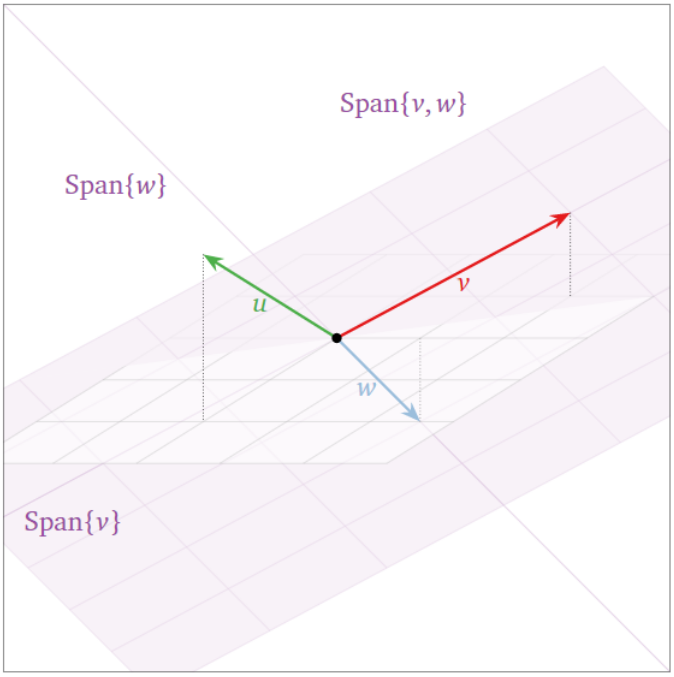


Figure 1.9.12

The three coplanar vectors $\{v, w, u\}$ below are linearly dependent:

- u is in $\text{Span}\{v, w\}$, so we can apply the first criterion, Theorem 1.9.1.
- The span did not increase when we added u, so we can apply the increasing span criterion, Theorem 1.9.2.

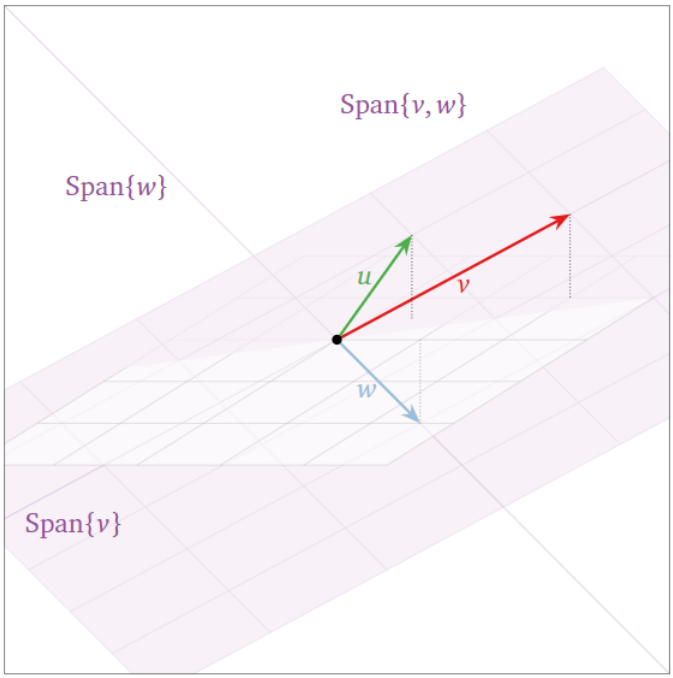


Figure 1.9.13

Note that three vectors are linearly dependent if and only if they are *coplanar*. Indeed, $\{v, w, u\}$ is linearly dependent if and only if one vector is in the span of the other two, which is a plane (or a line) (or $\{0\}$).

The four vectors $\{v, w, u, x\}$ below are linearly dependent: they are the columns of a wide matrix, see Note 1.9.2. Note however that u is not contained in $\text{Span}\{v, w, x\}$. See this warning, Note 1.9.3.

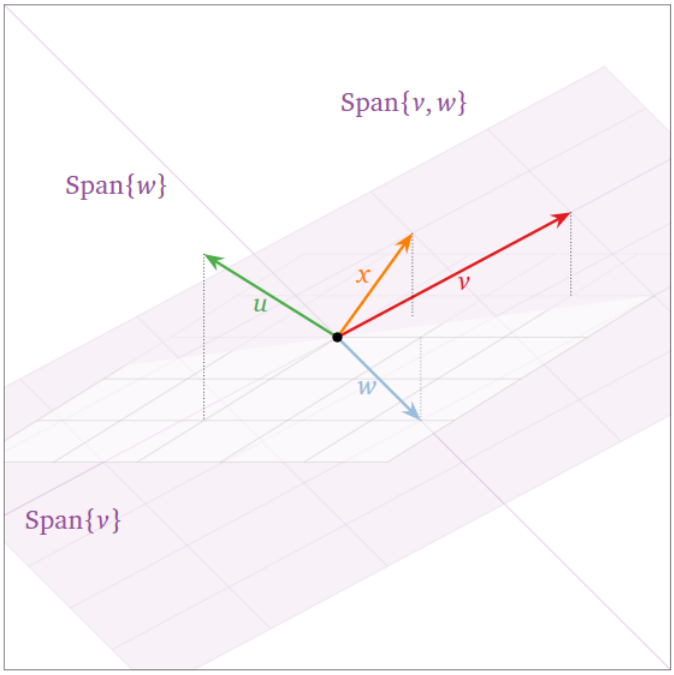


Figure 1.9.14: The vectors $\{v, w, u, x\}$ are linearly dependent, but u is not contained in $\text{Span}\{v, w, x\}$.

Example 1.9.6: Interactive: Linear independence of two vectors in $\mathbb{R}^3$

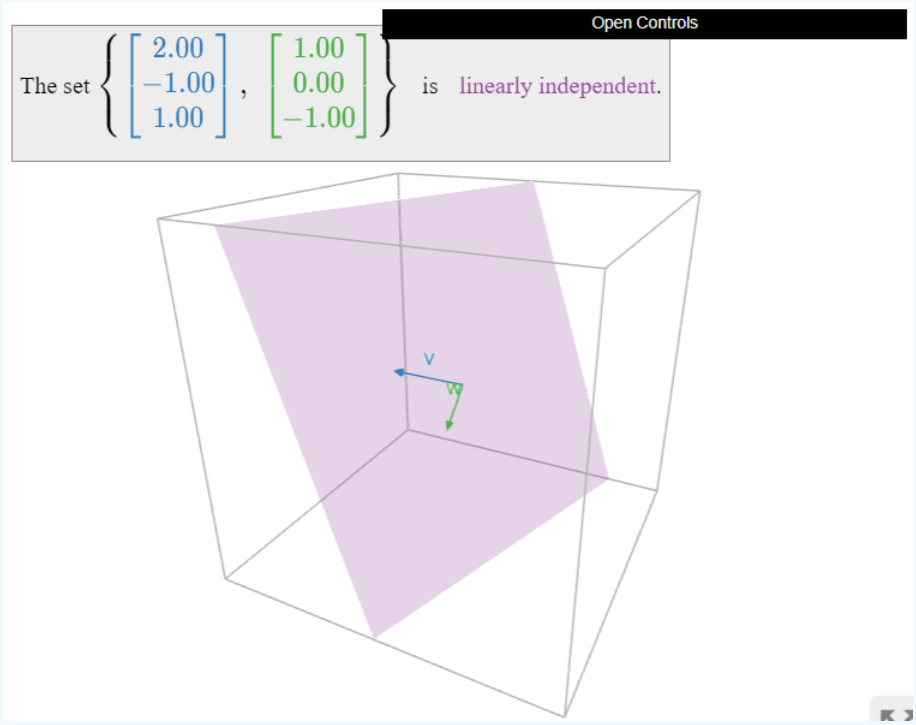


Figure 1.9.15: Move the vector heads and the demo will tell you if they are linearly independent and show you their span.

Example 1.9.7: Interactive: Linear independence of three vectors in $\mathbb{R}^3$

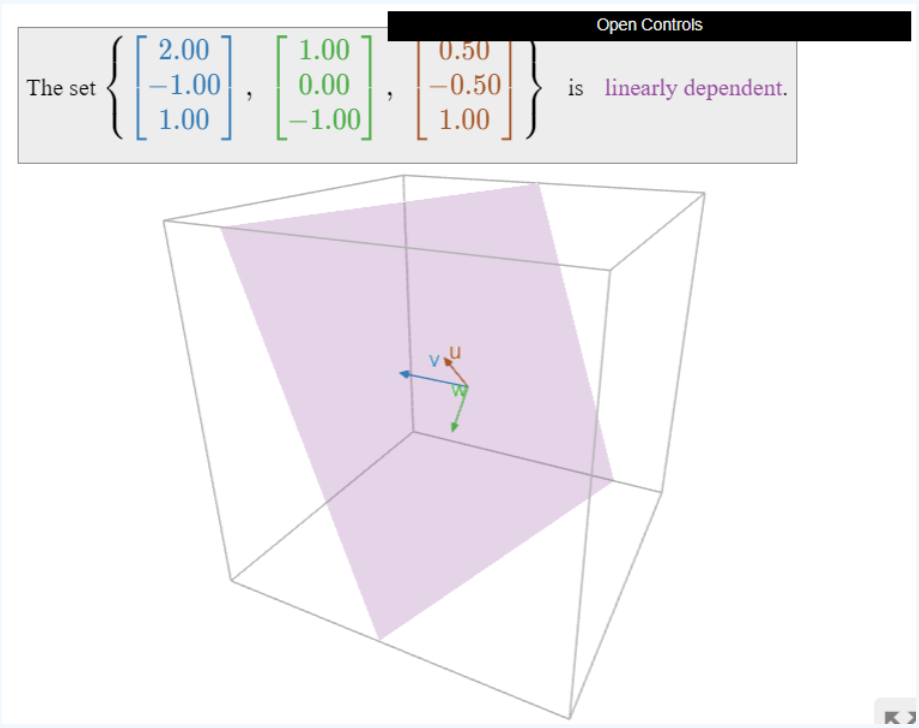


Figure 1.9.16: Move the vector heads and the demo will tell you if they are linearly independent and show you their span.

Linear Dependence and Free Variables

In light of this important note, Recipe: Checking Linear Independence, and this criterion, Theorem 1.9.1, it is natural to ask which columns of a matrix are redundant, i.e., which we can remove without affecting the column span.

Theorem 1.9.3

Let $v_1, v_2, \ldots, v_k$ be vectors in $\mathbb{R}^n$, and consider the matrix

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & & | \end{pmatrix}.$$

Then we can delete the columns of A *without* pivots (the columns corresponding to the free variables), without changing $\text{Span}\{v_1, v_2, \ldots, v_k\}$.

The pivot columns are linearly independent, so we cannot delete any more columns without changing the span.

Proof

If the matrix is in reduced row echelon form:

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

then the column without a pivot is visibly in the span of the pivot columns:

$$\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 0 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

and the pivot columns are linearly independent:

$$\begin{pmatrix}0\\0\\0\end{pmatrix} = x_1\begin{pmatrix}1\\0\\0\end{pmatrix} + x_2\begin{pmatrix}0\\1\\0\end{pmatrix} + x_4\begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}x_1\\x_2\\x_4\end{pmatrix} \implies x_1 = x_2 = x_4 = 0.$$

If the matrix is not in reduced row echelon form, then we row reduce:

$$A = \begin{pmatrix}1 & 7 & 23 & 3\\2 & 4 & 16 & 0\\-1 & -2 & -8 & 4\end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix}1 & 0 & 2 & 0\\0 & 1 & 3 & 0\\0 & 0 & 0 & 1\end{pmatrix}.$$

The following two vector equations have the same solution set, as they come from row-equivalent matrices:

$$x_1\begin{pmatrix}1\\2\\-1\end{pmatrix} + x_2\begin{pmatrix}7\\4\\-2\end{pmatrix} + x_3\begin{pmatrix}23\\16\\-8\end{pmatrix} + x_4\begin{pmatrix}3\\0\\4\end{pmatrix} = 0$$

$$x_1\begin{pmatrix}1\\0\\0\end{pmatrix} + x_2\begin{pmatrix}0\\1\\0\end{pmatrix} + x_3\begin{pmatrix}2\\3\\0\end{pmatrix} + x_4\begin{pmatrix}0\\0\\1\end{pmatrix} = 0$$

We conclude that

$$\begin{pmatrix}23\\16\\-8\end{pmatrix} = 2\begin{pmatrix}1\\2\\-1\end{pmatrix} + 3\begin{pmatrix}7\\4\\-2\end{pmatrix} + 0\begin{pmatrix}3\\0\\4\end{pmatrix}$$

and that

$$x_1\begin{pmatrix}1\\2\\-1\end{pmatrix} + x_2\begin{pmatrix}7\\4\\-2\end{pmatrix} + x_4\begin{pmatrix}3\\0\\4\end{pmatrix} = 0$$

has only the trivial solution.

Note that it is necessary to row reduce A to find which are its pivot columns, Definition 1.2.5 in Section 1.2. However, the span of the columns of the row reduced matrix is generally *not* equal to the span of the columns of A: one must use the pivot columns of the *original* matrix. See theorem in Section 2.7, Theorem 2.7.2 for a restatement of the above theorem.

Example 1.9.8

The matrix

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

has reduced row echelon form

$$\begin{pmatrix}1 & 0 & -8 & -7\\0 & 1 & 4 & 3\\0 & 0 & 0 & 0\end{pmatrix}.$$

Therefore, the first two columns of A are the pivot columns, so we can delete the others without changing the span:

$$\text{Span}\left\{\begin{pmatrix}1\\-2\\2\end{pmatrix}, \begin{pmatrix}2\\-3\\4\end{pmatrix}\right\} = \text{Span}\left\{\begin{pmatrix}1\\-2\\2\end{pmatrix}, \begin{pmatrix}2\\-3\\4\end{pmatrix}, \begin{pmatrix}0\\4\\0\end{pmatrix}, \begin{pmatrix}-1\\5\\-2\end{pmatrix}\right\}.$$

Moreover, the first two columns are linearly independent.

Note 1.9.5: Pivot Columns and Dimension

Let d be the number of pivot columns in the matrix

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & & | \end{pmatrix}.$$

- If $d = 1$ then $\text{Span}\{v_1, v_2, \ldots, v_k\}$ is a line.
- If $d = 2$ then $\text{Span}\{v_1, v_2, \ldots, v_k\}$ is a plane.
- If $d = 3$ then $\text{Span}\{v_1, v_2, \ldots, v_k\}$ is a 3-space.
- Et cetera.

The number d is called the dimension. We discussed this notion in this important note in Section 2.4, Note 2.4.4 and this important note in Section 2.4, Note 2.4.5. We will define this concept rigorously in Section 2.7.

This page titled 1.9: Linear Independence is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **2.5: Linear Independence** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

1.10: Exercises

Exercises

Exercise 1.10.1

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.

Answer

$\begin{matrix} x+3y=1 \\ 4x-y=3 \end{matrix}$, Solution is: $\left[x=\frac{10}{13}, y=\frac{1}{13}\right]$.

Exercise 1.10.2

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.

Answer

$\begin{matrix} 3x+y=3 \\ x+2y=1 \end{matrix}$, Solution is: $[x=1, y=0]$

Exercise 1.10.3

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.

Answer

$\begin{matrix} x+2y=1 \\ 2x-y=1 \\ 4x+3y=3 \end{matrix}$, Solution is: $\left[x=\frac{3}{5}, y=\frac{1}{5}\right]$

Exercise 1.10.4

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.

Answer

No solution exists. You can see this by writing the augmented matrix and doing row operations. $\begin{bmatrix} 1 & 1 & -3 & 2 \\ 2 & 1 & 1 & 1 \\ 3 & 2 & -2 & 0 \end{bmatrix}$, row echelon form: $\begin{bmatrix} 1 & 0 & 4 & 0 \\ 0 & 1 & -7 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Thus one of the equation says $0=1$ in an equivalent system of equations.

Exercise 1.10.5

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.

Answer

$$\begin{matrix} 4g-I=150 \\ 4I-17g=-660 \\ 4g+s=290 \\ g+I+s-b=0 \end{matrix}$$, Solution is: $\{g=60, I=90, b=200, s=50\}$

Exercise 1.10.6

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

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

Answer

The solution exists but is not unique.

Exercise 1.10.7

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

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

Answer

A solution exists and is unique.

Exercise 1.10.8

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

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

Exercise 1.10.9

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

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

Answer

There might be a solution. If so, there are infinitely many.

Exercise 1.10.10

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.

Answer

No. Consider $x+y+z=2$ and $x+y+z=1$.

Exercise 1.10.11

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.

Answer

These can have a solution. For example, $x+y=1, 2x+2y=2, 3x+3y=3$ even has an infinite set of solutions.

Exercise 1.10.12

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.

Answer

$h=4$

Exercise 1.10.13

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.

Answer

Any h will work.

Exercise 1.10.14

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.

Answer

Any h will work.

Exercise 1.10.15

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

a. one solution
b. no solution
c. infinitely many solutions

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

Answer

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.

Exercise 1.10.16

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

a. one solution
b. no solution
c. infinitely many solutions

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

Answer

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.

Exercise 1.10.17

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

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

Answer

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

Exercise 1.10.18

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

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

Answer

Solution is: $\left[w=\frac{3}{2}y-1, x=\frac{2}{3}-\frac{1}{2}y, z=\frac{1}{3}\right]$

Exercise 1.10.19

Determine which matrices are in reduced row-echelon form.

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

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

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

Answer

a. This one is not.
b. This one is.
c. This one is.

Exercise 1.10.20

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 1.10.21

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 1.10.22

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

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

Exercise 1.10.23

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

$$\begin{bmatrix} 2 & 4 & 5 & 15 \\ 1 & 2 & 3 & 9 \\ 1 & 2 & 2 & 6 \end{bmatrix}$$

Exercise 1.10.24

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

$$\begin{bmatrix} 4 & -1 & 7 & 10 \\ 1 & 0 & 3 & 3 \\ 1 & -1 & -2 & 1 \end{bmatrix}$$

Exercise 1.10.25

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

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

Exercise 1.10.26

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

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

Exercise 1.10.27

Find the solution of the system whose augmented matrix is

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

Exercise 1.10.28

Find the solution of the system whose augmented matrix is

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

Answer

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 $t \in \mathbb{R}$.

Exercise 1.10.29

Find the solution of the system whose augmented matrix is

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

Answer

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 $z = t, y = 4t, x = 2 - 4t$.

Exercise 1.10.30

Find the solution of the system whose augmented matrix is

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

Answer

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 $x_5 = t, x_4 = 1 - 6t, x_3 = -1 + 7t, x_2 = 4t, x_1 = 3 - 9t$.

Exercise 1.10.31

Find the solution of the system whose augmented matrix is

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

Answer

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 $x_4 = -\frac{1}{2} - \frac{3}{2}t, x_2 = \frac{3}{2} - t\frac{1}{2}, , x_1 = \frac{5}{2} + \frac{1}{2}t - 2s$.

Exercise 1.10.32

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

Answer

Solution is: $[x = 1 - 2t, z = 1, y = t]$

Exercise 1.10.33

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

Answer

Solution is: $[x=2-4t, y=-8t, z=t]$

Exercise 1.10.34

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

Answer

Solution is: $[x=-1, y=2, z=-1]$

Exercise 1.10.35

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

Answer

Solution is: $[x=2, y=4, z=5]$

Exercise 1.10.36

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

Answer

Solution is: $[x=1, y=2, z=-5]$

Exercise 1.10.37

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

Answer

Solution is: $[x=-1, y=-5, z=4]$

Exercise 1.10.38

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

Answer

Solution is: $[x=2t+1, y=4t, z=t]$

Exercise 1.10.39

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

Answer

Solution is: $[x=1, y=5, z=3]$

Exercise 1.10.40

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

Answer

Solution is: $[x=4, y=-4, z=-2]$

Exercise 1.10.41

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.

Answer

No. Consider $x+y+z=2$ and $x+y+z=1$.

Exercise 1.10.42

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.

Answer

No. This would lead to $0=1$.

Exercise 1.10.43

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.

Answer

Yes. It has a unique solution.

Exercise 1.10.44

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

Answer

The last column must not be a pivot column. The remaining columns must each be pivot columns.

Exercise 1.10.45

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 $z=0.05in\frac{1}{4}(10+0+w+x)$.

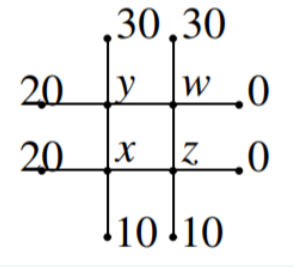


Figure 1.10.1

Answer

You need $\begin{matrix} \frac{1}{4}(20+30+w+x)-y=0 \\ \frac{1}{4}(y+30+0+z)-w=0 \\ \frac{1}{4}(20+y+z+10)-x=0 \\ \frac{1}{4}(x+w+0+10)-z=0 \end{matrix}$, Solution is: $[w=15, x=15, y=20, z=10]$.

? Exercise 1.10.46

Find the rank of the following matrix.

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

? Exercise 1.10.47

Find the rank of the following matrix.

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

? Exercise 1.10.48

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 1.10.49

Find the rank of the following matrix.

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

Exercise 1.10.50

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 1.10.51

Find the rank of the following matrix.

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

Exercise 1.10.52

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 1.10.53

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 1.10.54

Find the rank of the following matrix.

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

Exercise 1.10.55

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 1.10.56

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 1.10.57

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

Answer

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.

Exercise 1.10.58

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.

a. A is a 5×6 matrix, $rank(A) = 4$ and $rank[A|B] = 4$.
b. A is a 3×4 matrix, $rank(A) = 3$ and $rank[A|B] = 2$.
c. A is a 4×2 matrix, $rank(A) = 4$ and $rank[A|B] = 4$.
d. A is a 5×5 matrix, $rank(A) = 4$ and $rank[A|B] = 5$.
e. A is a 4×2 matrix, $rank(A) = 2$ and $rank[A|B] = 2$.

Answer

a. This says B is in the span of four of the columns. Thus the columns are not independent. Infinite solution set.
b. This surely can't happen. If you add in another column, the rank does not get smaller.
c. 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.
d. 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.
e. In this case, there is a unique solution since the columns of A are independent.

Exercise 1.10.59

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?

Answer

These are not legitimate row operations. They do not preserve the solution set of the system.

Exercise 1.10.60

Balance the following chemical reactions.

a. $KNO_3 + H_2CO_3 \rightarrow K_2CO_3 + HNO_3$
b. $AgI + Na_2S \rightarrow Ag_2S + NaI$

c. $Ba_3N_2 + H_2O \rightarrow Ba(OH)_2 + NH_3$
d. $CaCl_2 + Na_3PO_4 \rightarrow Ca_3(PO_4)_2 + NaCl$

Exercise 1.10.61

In the section on dimensionless variables it was observed that $\rho V^2 AB$ has the units of force. Describe a systematic way to obtain such combinations of the variables which will yield something which has the units of force.

Exercise 1.10.62

Consider the following diagram of four circuits.

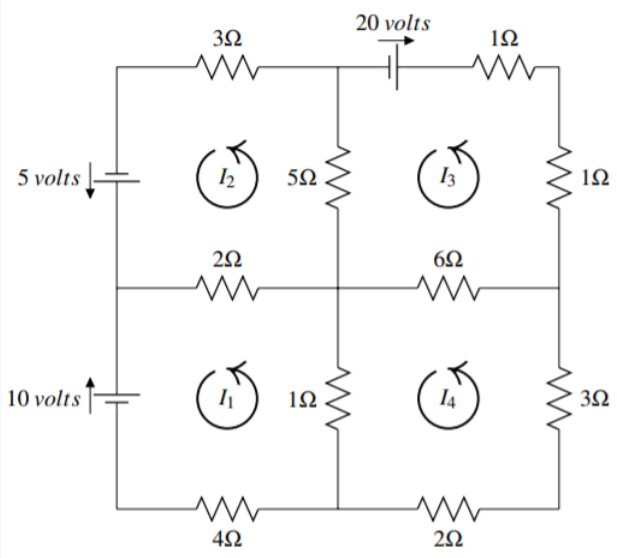


Figure 1.10.2

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.

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.

Answer

The other two equations are

$$\begin{aligned} 6I_3 - 6I_4 + I_3 + I_3 + 5I_3 - 5I_2 &= -20 \\ 2I_4 + 3I_4 + 6I_4 - 6I_3 + I_4 - I_1 &= 0 \end{aligned}$$

Then the system is

$$\begin{aligned} 2I_2 - 2I_1 + 5I_2 - 5I_3 + 3I_2 &= 5 \\ 4I_1 + I_1 - I_4 + 2I_1 - 2I_2 &= -10 \\ 6I_3 - 6I_4 + I_3 + I_3 + 5I_3 - 5I_2 &= -20 \\ 2I_4 + 3I_4 + 6I_4 - 6I_3 + I_4 - I_1 &= 0 \end{aligned}$$

The solution is:

$$I_1 = -\frac{750}{373}$$
$$I_2 = -\frac{1421}{1119}$$
$$I_3 = -\frac{3061}{1119}$$
$$I_4 = -\frac{1718}{1119}$$

? Exercise 1.10.63

Consider the following diagram of three circuits.

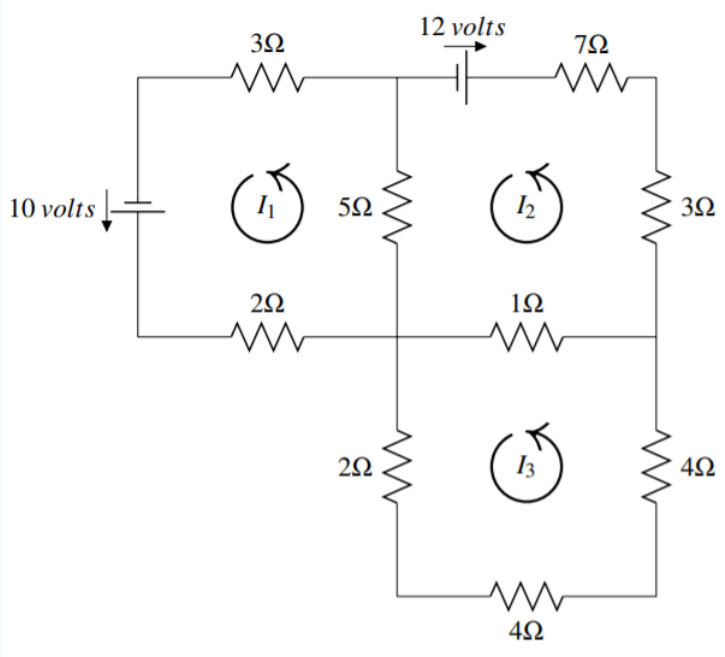


Figure 1.10.3

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.

Find I_1, I_2, I_3.

Answer

You have

$$\begin{aligned} 2I_1 + 5I_1 + 3I_1 - 5I_2 &= 10 \\ I_2 - I_3 + 3I_2 + 7I_2 + 5I_2 - 5I_1 &= -12 \\ 2I_3 + 4I_3 + 4I_3 + I_3 - I_2 &= 0 \end{aligned}$$

Simplifying this yields

$$\begin{aligned} 10I_1 - 5I_2 &= 10 \\ -5I_1 + 16I_2 - I_3 &= -12 \\ -I_2 + 11I_3 &= 0 \end{aligned}$$

The solution is given by

$$I_1 = \frac{218}{295}, I_2 = -\frac{154}{295}, I_3 = -\frac{14}{295}$$

This page titled 1.10: Exercises is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.9: Exercises** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

CHAPTER OVERVIEW

2: Matrices

2.1: Matrix Arithmetic
2.2: Multiplication of Matrices
2.3: The ijth Entry of a Product
2.4: Properties of Matrix Multiplication
2.5: The Transpose
2.6: The Identity and Inverses
2.7: Finding the Inverse of a Matrix
2.8: Elementary Matrices
2.9: More on Matrix Inverses
2.10: LU Factorization
2.E: Exercises

This page titled 2: Matrices is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

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.
D. Solve a linear system using matrix algebra.
E. Use multiplication by an elementary matrix to apply row operations.
F. 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.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{bmatrix} \tag{2.1.1}$$

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.

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

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix}$$

Therefore, A_j is the jth 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 $(\mathbf{i,j})$**-entry** of a matrix is the entry in the ith row and j^{th} column. For example, in the matrix (2.1.1) 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 ith row and the jth 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 (2.1.1), $a_{23} = 8, a_{32} = -9, a_{12} = 2,$ 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.

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 2.1.1: The Zero Matrix

The 2×3 zero matrix is $0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

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

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.

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.

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.

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 2.1.4, 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 2.1.1: Properties of Matrix Addition

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

- Commutative Law of Addition

$$A+B = B+A \tag{2.1.2}$$

- Associative Law of Addition

$$(A+B)+C = A+(B+C) \tag{2.1.3}$$

- Existence of an Additive Identity

$$\begin{gathered} \text{There exists a zero matrix } 0 \text{ such that} \\ A+0 = A \end{gathered} \tag{2.1.4}$$

- Existence of an Additive Inverse

$$\begin{gathered} \text{There exists a matrix } -A \text{ such that} \\ A+(-A) = 0 \end{gathered} \tag{2.1.5}$$

Proof

Consider the Commutative Law of Addition given in (2.1.2). 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 2.1.4 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 (2.1.4) the **additive identity**. Similarly, we call the matrix $-A$ in (2.1.5) 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 $(-1)A$.

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.

$$3\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 6 & 9 & 12 \\ 15 & 6 & 24 & 21 \\ 18 & -27 & 3 & 6 \end{bmatrix}$$

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 2.1.5: Scalar Multiplication of Matrices

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

Consider the following example.

Example 2.1.3: Effect of Multiplication by a Scalar

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

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

Solution

By Definition 2.1.5, we multiply each element of A by 7. Therefore,

$$7A = 7\begin{bmatrix} 2 & 0 \\ 1 & -4 \end{bmatrix} = \begin{bmatrix} 7(2) & 7(0) \\ 7(1) & 7(-4) \end{bmatrix} = \begin{bmatrix} 14 & 0 \\ 7 & -28 \end{bmatrix}$$

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

Proposition 2.1.2: 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

$$1A = A$$

Proof

The proof of this proposition is similar to the proof of Proposition 2.1.1and is left an exercise to the reader.

This page titled 2.1: Matrix Arithmetic is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.1: Matrix Arithmetic** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

2.2: 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 2.2.1: 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 ith row of a column matrix, or the ith column of a row matrix.

The $n \times 1$ matrix

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

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

$$X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$$

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 9.2.2. In this context, a linear combination is a sum consisting of vectors multiplied by scalars. For example,

$$\begin{bmatrix} 50 \\ 122 \end{bmatrix} = 7 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + 8 \begin{bmatrix} 2 \\ 5 \end{bmatrix} + 9 \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

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 2.2.2: The Vector Form of a System of Linear Equations

Suppose we have a system of equations given by

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

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

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

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

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

We will soon see that this equals

$$7 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + 8 \begin{bmatrix} 2 \\ 5 \end{bmatrix} + 9 \begin{bmatrix} 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 50 \\ 122 \end{bmatrix}$$

In general terms,

$$\begin{aligned} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &= x_1 \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} + x_3 \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \end{bmatrix} \end{aligned}$$

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 2.2.3: 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 2.2.1: 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 2.2.3 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 2.2.4: The Matrix Form of a System of Linear Equations

Suppose we have a system of equations given by

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

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 2.2.1, 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

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

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 2.2.5: 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 $B_1, \ldots, B_p$ 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 kth column of AB.

Consider the following example.

Example 2.2.2: Multiplying Two Matrices

Find AB if possible.

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

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 2.2.5 gives us a way to calculate each column of AB, as follows.

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

You know how to multiply a matrix times a vector, using Definition 2.2.3 for each of the three columns. Thus

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 9 & 3 \\ -2 & 7 & 3 \end{bmatrix}$$

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

Example 2.2.3: Vector Times Vector Multiplication

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

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 2.2.5, we can compute this product as follows

$$\begin{bmatrix}1\\2\\1\end{bmatrix}\begin{bmatrix}1 & 2 & 1 & 0\end{bmatrix}=\begin{bmatrix}\overbrace{\begin{bmatrix}1\\2\\1\end{bmatrix}\begin{bmatrix}1\end{bmatrix}}^{\text{First column}}, & \overbrace{\begin{bmatrix}1\\2\\1\end{bmatrix}\begin{bmatrix}2\end{bmatrix}}^{\text{Second column}}, & \overbrace{\begin{bmatrix}1\\2\\1\end{bmatrix}\begin{bmatrix}1\end{bmatrix}}^{\text{Third column}}, & \overbrace{\begin{bmatrix}1\\2\\1\end{bmatrix}\begin{bmatrix}0\end{bmatrix}}^{\text{Fourth column}}\end{bmatrix}$$

You can use Definition 2.2.3 to verify that this product is

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

Example 2.2.4: A Multiplication Which is Not Defined

Find BA if possible.

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

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 2.2.2. 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 2.2.5: Multiplication by the Zero Matrix

Compute the product $A0$ for the matrix

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

and the 2×2 zero matrix given by

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

Solution

In this product, we compute

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

Hence, $A0 = 0$.

Notice that we could also multiply A by the 2×1 zero vector given by $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$. 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.

This page titled 2.2: Multiplication of Matrices is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.2: Multiplication of Matrices** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

2.3: The ijth 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 ith row and the jth 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 jth 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 2.3.1: 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

$$(AB)_{ij} = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

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

Consider the following example.

Example 2.3.1: The Entries of a Product

Compute AB if possible. If it is, find the $(3,2)$-entry of AB using Definition 2.3.1.

$$A=\begin{bmatrix}1&2\\3&1\\2&6\end{bmatrix}, B=\begin{bmatrix}2&3&1\\7&6&2\end{bmatrix}$$

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:

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

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

$$\begin{bmatrix}16&15&5\\13&15&5\\46&42&14\end{bmatrix}$$

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

Now using Definition 2.3.1, 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 2.3.2: Finding the Entries of a Product

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

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

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 2.3.1, 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

$$(AB)_{21}=\sum_{k=1}^{3} a_{2k}b_{k1}=\begin{bmatrix}a_{21}&a_{22}&a_{23}\end{bmatrix}\begin{bmatrix}b_{11}\\b_{21}\\b_{31}\end{bmatrix}$$

Substituting in the appropriate values, this product becomes

$$\begin{bmatrix} a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} = \begin{bmatrix} 7 & 6 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = 1 \times 7 + 3 \times 6 + 2 \times 2 = 29$$

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

$$AB = \begin{bmatrix} 13 & 13 \\ 29 & 32 \\ 0 & 0 \end{bmatrix}$$

This page titled 2.3: The ijth Entry of a Product is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.3: The ijth Entry of a Product** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

2.4: 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 2.4.1: Matrix Multiplication is Not Commutative

Compare the products AB and BA, for matrices $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

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

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

The second product, BA is

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

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 2.4.1: Properties of Matrix Multiplication

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

$$A(rB + sC) = r(AB) + s(AC) \tag{2.4.1}$$

$$(B + C)A = BA + CA \tag{2.4.2}$$

$$A(BC) = (AB)C \tag{2.4.3}$$

Proof

First we will prove (2.4.1). We will use Definition 2.3.1 and prove this statement using the ij^{th} entries of a matrix. Therefore,

$$\begin{aligned} (A(rB + sC))_{ij} &= \sum_k a_{ik} (rB + sC)_{kj} \\ &= \sum_k a_{ik} (rb_{kj} + sc_{kj}) \\ &= r \sum_k a_{ik} b_{kj} + s \sum_k a_{ik} c_{kj} \\ &= r(AB)_{ij} + s(AC)_{ij} \\ &= (r(AB) + s(AC))_{ij} \end{aligned}$$

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

The proof of Equation (2.4.2) follows the same pattern and is left as an exercise.

Statement Equation (2.4.3) is the **associative law of multiplication**. Using Definition 2.3.1,

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

This proves (2.4.3).

This page titled 2.4: Properties of Matrix Multiplication is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.4: Properties of Matrix Multiplication** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

2.5: 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.

$$\begin{bmatrix} 1 & 4 \\ 3 & 1 \\ 2 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 & 2 \\ 4 & 1 & 6 \end{bmatrix}$$

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 2.5.1: 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 2.5.1: The Transpose of a Matrix

Calculate A^T for the following matrix

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

Solution

By Definition 2.5.1, 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,

$$A^T = \begin{bmatrix} 1 & 3 \\ 2 & 5 \\ -6 & 4 \end{bmatrix}$$

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 2.5.1: 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 2.5.1,

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

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 2.5.2: 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 2.5.2: Symmetric Matrices

Let

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

Use Definition 2.5.2 to show that A is symmetric.

Solution

By Definition 2.5.2, we need to show that $A = A^T$. Now, using Definition 2.5.1,

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

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

Example 2.5.3: 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 2.5.2,

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

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 2.5.2, A is skew symmetric.

This page titled 2.5: The Transpose is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.5: The Transpose** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

2.6: 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 2.6.1: 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.

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

Definition 2.6.1: 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 2.6.1: 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.

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

Example 2.6.1: Verifying the Inverse of a Matrix

Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$. Show $\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$ 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 2.6.2: 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.

This page titled 2.6: The Identity and Inverses is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.6: The Identity and Inverses** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

2.7: Finding the Inverse of a Matrix

In Example 2.6.1, 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

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

as in Example 2.6.1. In order to find A^{-1}, we need to find a matrix $\begin{bmatrix} x & z \\ y & w \end{bmatrix}$ such that

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

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,

$$\begin{aligned} x + y &= 1 \\ x + 2y &= 0 \end{aligned}$$

and

$$\begin{aligned} z + w &= 0 \\ z + 2w &= 1 \end{aligned}$$

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

$$\left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 2 & 1 \end{array}\right] \tag{2.7.1}$$

for the second.

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

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

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

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

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

Now solve the second system, (2.7.1) 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

$$A^{-1} = \begin{bmatrix} x & z \\ y & w \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

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,

$$\begin{bmatrix} x \\ y \end{bmatrix}$$

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

$$\begin{bmatrix} z \\ w \end{bmatrix}$$

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

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

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

This exploration motivates the following important algorithm.

Algorithm 2.7.1: 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 2.7.1: Finding the Inverse

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

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

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 2.7.2: A Matrix Which Has No Inverse

Let $A = \begin{bmatrix} 1 & 2 & 2 \\ 1 & 0 & 2 \\ 2 & 2 & 4 \end{bmatrix}$. Find A^{-1} if it exists.

Solution

Write the augmented matrix $[A|I]$

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

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.

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

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

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

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.

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 2.6.1. 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 2.2.4 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 2.7.3: 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

$$AX = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = B \tag{2.7.2}$$

The inverse of the matrix

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

is

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

Verifying this inverse is left as an exercise.

From here, the solution to the given system (2.7.2) is found by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = A^{-1}B = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{5}{2} \\ -2 \\ -\frac{3}{2} \end{bmatrix}$$

What if the right side, B, of (2.7.2) had been $\begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$? In other words, what would be the solution to

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}?$$

By the above discussion, the solution is given by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = A^{-1}B = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix}$$

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 2.7.1: Inverses of Transposes and Products

Let A, B, and A_i for $i = 1, \ldots, k$ 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 $A_1, A_2, \ldots, A_k$ are invertible, then the product $A_1 A_2 \cdots A_k$ is invertible, and $(A_1 A_2 \cdots A_k)^{-1} = A_k^{-1} A_{k-1}^{-1} \cdots A_2^{-1} A_1^{-1}$

Consider the following theorem.

Theorem 2.7.2: 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 $(pA)^{-1} = \frac{1}{p}A^{-1}$

This page titled 2.7: Finding the Inverse of a Matrix is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.7: Finding the Inverse of a Matrix** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

2.8: 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 1.3.2. 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

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

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 2.8.1: 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 2.8.1: 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 1.3.2.

First, consider the following lemma.

Lemma 2.8.1: Action of Permutation Matrix

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

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

where B is obtained from A by switching the ith and the jth rows.

We will explore this idea more in the following example.

Example 2.8.1: Switching Rows with an Elementary Matrix

Let

$$P^{12} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A = \begin{bmatrix} a & b \\ g & d \\ e & f \end{bmatrix}$$

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

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

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 2.8.1 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 2.8.2: Multiplication by a Scalar and Elementary Matrices

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

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

where B is obtained from A by multiplying the ith row of A by k.

We will explore this lemma further in the following example.

Example 2.8.2: 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 2.8.3: 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 ith row to the jth. Then

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

where B is obtained from A by adding k times the ith row to the jth row of A.

Consider the following example.

Example 2.8.3: 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 2.8.2: 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 2.8.4: Inverse of an Elementary Matrix

Let

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

Find E^{-1}.

Solution

Consider the elementary matrix E given by

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

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

$$E^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

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

$$EE^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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 2.8.2: 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 2.8.5: The Form $B = UA$

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

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

The resulting matrix is equivalent to finding the product of $P^{12} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and A.

Next, add (-2) times row 1 to row 3.

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

This is equivalent to multiplying by the matrix $E(-2\times 1+3)=\begin{bmatrix}1&0&0\\0&1&0\\-2&0&1\end{bmatrix}$. Notice that the resulting matrix is B, the required reduced row-echelon form of A.

We can then write

$$\begin{aligned}B&=E(-2\times 1+2)\left(P^{12}A\right)\\&=\left(E(-2\times 1+2)P^{12}\right)A\\&=UA\end{aligned}$$

It remains to find the matrix U.

$$\begin{aligned}U&=E(-2\times 1+2)P^{12}\\&=\begin{bmatrix}1&0&0\\0&1&0\\-2&0&1\end{bmatrix}\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}\\&=\begin{bmatrix}0&1&0\\1&0&0\\0&-2&1\end{bmatrix}\end{aligned}$$

We can verify that $B=UA$ holds for this matrix U:

$$\begin{aligned}UA&=\begin{bmatrix}0&1&0\\1&0&0\\0&-2&1\end{bmatrix}\begin{bmatrix}0&1\\1&0\\2&0\end{bmatrix}\\&=\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}\\&=B\end{aligned}$$

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 2.8.3: 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 2.8.3.

Example 2.8.6: The Form $B=UA$, Revisited

Let $A=\begin{bmatrix}0&1\\1&0\\2&0\end{bmatrix}$. Using the process outlined in Theorem 2.8.3, find U such that $B=UA$.

Solution

First, set up the matrix $[A|I_m]$.

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

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 2.8.5, you can see that the same matrix is obtained regardless of which process is used.

Recall from Algorithm 2.7.1 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 2.8.4: 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 2.8.7: Product of Elementary Matrices

Let $A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}$. Write A as a product of elementary matrices.

Solution

We will use the process outlined in Theorem 2.8.3 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:

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

represented by the elementary matrix $E_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Secondly:

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

represented by the elementary matrix $E_2 = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Finally:

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

represented by the elementary matrix $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}$.

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.

$$\begin{aligned} U^{-1} &= (E_3 E_2 E_1)^{-1} \\ &= E_1^{-1} E_2^{-1} E_3^{-1} \\ &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \\ &= A \end{aligned}$$

This gives A written as a product of elementary matrices. By Theorem 2.8.4 it follows that A is invertible.

This page titled 2.8: Elementary Matrices is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.8: Elementary Matrices** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

2.9: 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 2.9.1: 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 2.9.2: 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 2.9.1, 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 2.9.1: 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 2.9.2 with A and B, to obtain the inequality $m \leq n$. We then apply Lemma 2.9.2 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 2.9.1: 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 2.9.2: 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 2.9.1 gives us that R does not have a row of zeros. By noting that R is a square matrix and applying Proposition 2.9.1, 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 2.9.1: 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

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2.9.1}$$

Therefore, $A^T A = I_2$, where I_2 is the 2×2 identity matrix. However, the product AA^T is

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2.9.2}$$

Hence AA^T is not the 3×3 identity matrix. This shows that for Theorem 2.9.2, 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 2.9.3: 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 2.9.1, we get that the of A does not have a row of zeros. Then by Theorem 2.9.1, it follows that A and the reduced row-echelon form of A are square matrices. Finally, by Proposition 2.9.1, 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 2.9.2, we can conclude that A is invertible.

Theorem 2.9.3 corresponds to Algorithm 2.7.1, which claims that A^{-1} is found by row reducing the augmented matrix $[A|I]$ to the form $\left[I|A^{-1}\right]$. 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 2.9.3, if $EA \neq I$, then A is not invertible, and if $EA = I$, A is invertible. If $EA = I$, then by Theorem 2.9.2, $E = A^{-1}$. This proves that Algorithm 2.7.1 does in fact find A^{-1}.

This page titled 2.9: More on Matrix Inverses is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.9: More on Matrix Inverses** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

2.10: LU Factorization

An LU factorization of a matrix involves writing the given matrix as the product of a lower triangular matrix L which has the main diagonal consisting entirely of ones, and an upper triangular matrix U in the indicated order. This is the version discussed here but it is sometimes the case that the L has numbers other than 1 down the main diagonal. It is still a useful concept. The L goes with "lower" and the U with "upper".

It turns out many matrices can be written in this way and when this is possible, people get excited about slick ways of solving the system of equations, $AX = B$. It is for this reason that you want to study the LU factorization. It allows you to work only with triangular matrices. It turns out that it takes about half as many operations to obtain an LU factorization as it does to find the row reduced echelon form.

First it should be noted not all matrices have an LU factorization and so we will emphasize the techniques for achieving it rather than formal proofs.

Example 2.10.1: A Matrix with NO LU factorization

Can you write $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ in the form LU as just described?

Solution

To do so you would need

$$\begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} = \begin{bmatrix} a & b \\ xa & xb+c \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Therefore, $b = 1$ and $a = 0$. Also, from the bottom rows, $xa = 1$ which can't happen and have $a = 0$. Therefore, you can't write this matrix in the form It has no LU factorization. This is what we mean above by saying the method lacks generality.

Nevertheless the method is often extremely useful, and we will describe below one the many methods used to produce an LU factorization when possible.

Finding An LU Factorization By Inspection

Which matrices have an LU factorization? It turns out it is those whose row-echelon form can be achieved without switching rows. In other words matrices which only involve using row operations of type 2 or 3 to obtain the row-echelon form.

Example 2.10.2: An LU factorization

Find an LU factorization of $A = \begin{bmatrix} 1 & 2 & 0 & 2 \\ 1 & 3 & 2 & 1 \\ 2 & 3 & 4 & 0 \end{bmatrix}$.

Solution

One way to find the LU factorization is to simply look for it directly. You need

$$\begin{bmatrix} 1 & 2 & 0 & 2 \\ 1 & 3 & 2 & 1 \\ 2 & 3 & 4 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ x & 1 & 0 \\ y & z & 1 \end{bmatrix} \begin{bmatrix} a & d & h & j \\ 0 & b & e & i \\ 0 & 0 & c & f \end{bmatrix}.$$

Then multiplying these you get

$$\begin{bmatrix} a & d & h & j \\ xa & xd+b & xh+e & xj+i \\ ya & yd+zb & yh+ze+c & yj+iz+f \end{bmatrix}$$

and so you can now tell what the various quantities equal. From the first column, you need $a = 1, x = 1, y = 2$. Now go to the second column. You need $d = 2, xd+b = 3$ so $b = 1, yd+zb = 3$ so $z = -1$. From the third column, $h = 0, e = 2, c = 6$. Now from the fourth column, $j = 2, i = -1, f = -5$. Therefore, an LU factorization is

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

You can check whether you got it right by simply multiplying these two.

LU Factorization, Multiplier Method

Remember that for a matrix A to be written in the form $A = LU$, you must be able to reduce it to its row-echelon form without interchanging rows. The following method gives a process for calculating the LU factorization of such a matrix A.

Example 2.10.3: LU factorization

Find an LU factorization for

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

Solution

Write the matrix as the following product.

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

In the matrix on the right, begin with the left row and zero out the entries below the top using the row operation which involves adding a multiple of a row to another row. You do this and also update the matrix on the left so that the product will be unchanged. Here is the first step. Take -2 times the top row and add to the second. Then take 2 times the top row and add to the second in the matrix on the left.

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

The next step is to take 2 times the top row and add to the bottom in the matrix on the right. To ensure that the product is unchanged, you place a in the bottom left in the matrix on the left. Thus the next step yields

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

Next take 7 times the middle row on right and add to bottom row. Updating the matrix on the left in a similar manner to what was done earlier,

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -2 & -7 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -5 \\ 0 & 0 & -31 \end{bmatrix}$$

At this point, stop. You are done.

The method just described is called the multiplier method.

Solving Systems using LU Factorization

One reason people care about the LU factorization is it allows the quick solution of systems of equations. Here is an example.

Example 2.10.4: LU factorization to Solve Equations

Suppose you want to find the solutions to

$$\begin{bmatrix} 1 & 2 & 3 & 2 \\ 4 & 3 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Solution

Of course one way is to write the augmented matrix and grind away. However, this involves more row operations than the computation of the LU factorization and it turns out that the LU factorization can give the solution quickly. Here is how. The following is an LU factorization for the matrix.

$$\begin{bmatrix} 1 & 2 & 3 & 2 \\ 4 & 3 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{bmatrix}.$$

Let $UX = Y$ and consider $LY = B$ where in this case, $B = [1, 2, 3]^T$. Thus

$$\begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

which yields very quickly that $Y = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$.

Now you can find X by solving $UX = Y$. Thus in this case,

$$\begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$$

which yields

$$X = \begin{bmatrix} -\frac{3}{5} + \frac{7}{5}t \\ \frac{9}{5} - \frac{11}{5}t \\ t \\ -1 \end{bmatrix}, \quad t \in \mathbb{R}.$$

Justification for the Multiplier Method

Why does the multiplier method work for finding the LU factorization? Suppose A is a matrix which has the property that the row-echelon form for A may be achieved without switching rows. Thus every row which is replaced using this row operation in obtaining the row-echelon form may be modified by using a row which is above it.

Lemma 2.10.1: Multiplier Method and Triangular Matrices

Let L be a lower (upper) triangular matrix $m \times m$ which has ones down the main diagonal. Then L^{-1} also is a lower (upper) triangular matrix which has ones down the main diagonal. In the case that L is of the form

$$L = \begin{bmatrix} 1 & & & \\ a_1 & 1 & & \\ \vdots & & \ddots & \\ a_n & & & 1 \end{bmatrix} \tag{2.10.1}$$

where all entries are zero except for the left column and main diagonal, it is also the case that L^{-1} is obtained from L by simply multiplying each entry below the main diagonal in L with -1. The same is true if the single nonzero column is in another position.

Proof

Consider the usual setup for finding the inverse $\begin{bmatrix} L & I \end{bmatrix}$. Then each row operation done to L to reduce to row reduced echelon form results in changing only the entries in I below the main diagonal. In the special case of L given in (2.10.1) or the single nonzero column is in another position, multiplication by -1 as described in the lemma clearly results in L^{-1}.

For a simple illustration of the last claim,

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

Now let A be an $m \times n$ matrix, say

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

and assume A can be row reduced to an upper triangular form using only row operation 3. Thus, in particular, $a_{11} \neq 0$. Multiply on the left by $E_1 =$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\frac{a_{21}}{a_{11}} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{a_{m1}}{a_{11}} & 0 & \cdots & 1 \end{bmatrix}$$

This is the product of elementary matrices which make modifications in the first column only. It is equivalent to taking $-a_{21}/a_{11}$ times the first row and adding to the second. Then taking $-a_{31}/a_{11}$ times the first row and adding to the third and so forth. The quotients in the first column of the above matrix are the multipliers. Thus the result is of the form

$$E_1 A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a'_{1n} \\ 0 & a'_{22} & \cdots & a'_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & a'_{m2} & \cdots & a'_{mn} \end{bmatrix}$$

By assumption, $a'_{22} \neq 0$ and so it is possible to use this entry to zero out all the entries below it in the matrix on the right by multiplication by a matrix of the form $E_2 = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & E \end{bmatrix}$ where E is an $[m-1] \times [m-1]$ matrix of the form

$$E = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\frac{a'_{32}}{a'_{22}} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{a'_{m2}}{a'_{22}} & 0 & \cdots & 1 \end{bmatrix}$$

Again, the entries in the first column below the 1 are the multipliers. Continuing this way, zeroing out the entries below the diagonal entries, finally leads to

$$E_{m-1} E_{n-2} \cdots E_1 A = U$$

where U is upper triangular. Each E_j has all ones down the main diagonal and is lower triangular. Now multiply both sides by the inverses of the E_j in the reverse order. This yields

$$A = E_1^{-1} E_2^{-1} \cdots E_{m-1}^{-1} U$$

By Lemma 2.10.1, this implies that the product of those E_j^{-1} is a lower triangular matrix having all ones down the main diagonal.

The above discussion and lemma gives the justification for the multiplier method. The expressions

$$\frac{-a_{21}}{a_{11}}, \frac{-a_{31}}{a_{11}}, \cdots, \frac{-a_{m1}}{a_{11}}$$

denoted respectively by $M_{21}, \cdots, M_{m1}$ to save notation which were obtained in building E_1 are the multipliers. Then according to the lemma, to find E_1^{-1} you simply write

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ -M_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -M_{m1} & 0 & \cdots & 1 \end{bmatrix}$$

Similar considerations apply to the other E_j^{-1}. Thus L is a product of the form

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ -M_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -M_{m1} & 0 & \cdots & 1 \end{bmatrix} \cdots \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & -M_{m[m-1]} & 1 \end{bmatrix}$$

each factor having at most one nonzero column, the position of which moves from left to right in scanning the above product of matrices from left to right. It follows from what we know about the effect of multiplying on the left by an elementary matrix that the above product is of the form

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -M_{21} & 1 & \cdots & 0 & 0 \\ \vdots & -M_{32} & \ddots & \vdots & \vdots \\ -M_{[m-1]1} & \vdots & \cdots & 1 & 0 \\ -M_{m1} & -M_{m2} & \cdots & -M_{m[m-1]} & 1 \end{bmatrix}$$

In words, beginning at the left column and moving toward the right, you simply insert, into the corresponding position in the identity matrix, -1 times the multiplier which was used to zero out an entry in that position below the main diagonal in A, while retaining the main diagonal which consists entirely of ones. This is L.

This page titled 2.10: LU Factorization is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.10: LU Factorization** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

2.E: Exercises

Exercise 2.E.1

For the following pairs of matrices, determine if the sum $A+B$ is defined. If so, find the sum.

a. $A=\begin{bmatrix}1&0\\0&1\end{bmatrix}, B=\begin{bmatrix}0&1\\1&0\end{bmatrix}$

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

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

Exercise 2.E.2

For each matrix A, find the matrix $-A$ such that $A+(-A)=0$.

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

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

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

Exercise 2.E.3

In the context of Proposition 2.1.1, describe $-A$ and 0.

Answer

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.

2.1.2: Scalar Multiplication of Matrices

Exercise 2.E.4

For each matrix A, find the product $(-2)A, 0A$, and $3A$.

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

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

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

Exercise 2.E.5

Using only the properties given in Proposition 2.1.1 and Proposition 2.1.2, show $-A$ is unique.

Answer

Suppose B also works. Then

$$-A=-A+(A+B)=(-A+A)+B=0+B=B$$

Exercise 2.E.6

Using only the properties given in Proposition 2.1.1 and Proposition 2.1.2, show 0 is unique.

Answer

Suppose $0'$ also works. Then $0' = 0' + 0 = 0$.

Exercise 2.E.7

Using only the properties given in Proposition 2.1.1 and Proposition 2.1.2 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.

Answer

$0A = (0+0)\,A = 0A + 0A$. Now add $-(0A)$ to both sides. Then $0 = 0A$.

Exercise 2.E.8

Using only the properties given in Proposition 2.1.1 and Proposition 2.1.2, as well as previous problems show $(-1)\,A = -A$.

Answer

$A + (-1)\,A = (1 + (-1))\,A = 0A = 0$. Therefore, from the uniqueness of the additive inverse proved in the above Problem 2.E.7, it follows that $-A = (-1)\,A$.

2.2

Exercise 2.E.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.

a. $-3A$
b. $3B - A$
c. AC
d. CB
e. AE
f. EA

Answer

a. $\begin{bmatrix} -3 & -6 & -9 \\ -6 & -3 & -21 \end{bmatrix}$
b. $\begin{bmatrix} 8 & -5 & 3 \\ -11 & 5 & -4 \end{bmatrix}$
c. Not possible
d. $\begin{bmatrix} -3 & 3 & 4 \\ 6 & -1 & 7 \end{bmatrix}$
e. Not possible
f. Not possible

Exercise 2.E.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.

a. $-3A$
b. $3B - A$
c. AC
d. CA
e. AE
f. EA
g. BE
h. DE

Answer

a. $\begin{bmatrix} -3 & -6 \\ -9 & -6 \\ -3 & 3 \end{bmatrix}$
b. Not possible.
c. $\begin{bmatrix} 11 & 2 \\ 13 & 6 \\ -4 & 2 \end{bmatrix}$
d. Not possible.
e. $\begin{bmatrix} 7 \\ 9 \\ -2 \end{bmatrix}$
f. Not possible.
g. Not possible.
h. $\begin{bmatrix} 2 \\ -5 \end{bmatrix}$

Exercise 2.E.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.

a. AB
b. BA
c. AC
d. CA
e. CB
f. BC

Answer

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

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

c. Not possible

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

e. $\begin{bmatrix} 8 & 1 & -3 \\ 7 & 6 & -6 \end{bmatrix}$

Exercise 2.E.12

Let $A = \begin{bmatrix} -1 & -1 \\ 3 & 3 \end{bmatrix}$. Find all 2×2 matrices, B such that $AB = 0$.

Answer

$$\begin{bmatrix} -1 & -1 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} -x - z & -w - y \\ 3x + 3z & 3w + 3y \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Solution is: $w = -y, x = -z$ so the matrices are of the form $\begin{bmatrix} x & y \\ -x & -y \end{bmatrix}$.

Exercise 2.E.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.

Answer

$$X^T Y = \begin{bmatrix} 0 & -1 & -2 \\ 0 & -1 & -2 \\ 0 & 1 & 2 \end{bmatrix}, XY^T = 1$$

Exercise 2.E.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?

Answer

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & k \end{bmatrix} = \begin{bmatrix} 7 & 2k + 2 \\ 15 & 4k + 6 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 2 \\ 3 & k \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 7 & 10 \\ 3k + 3 & 4k + 6 \end{bmatrix}$$

Thus you must have $\begin{matrix} 3k + 3 = 15 \\ 2k + 2 = 10 \end{matrix}$, Solution is: $[k = 4]$

Exercise 2.E.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?

Answer

$$\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}\begin{bmatrix}1 & 2\\ 1 & k\end{bmatrix}=\begin{bmatrix}3 & 2k+2\\ 7 & 4k+6\end{bmatrix}$$
$$\begin{bmatrix}1 & 2\\ 1 & k\end{bmatrix}\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}=\begin{bmatrix}7 & 10\\ 3k+1 & 4k+2\end{bmatrix}$$

However, $7\neq 3$ and so there is no possible choice of k which will make these matrices commute.

Exercise 2.E.16

Find 2×2 matrices, A, B, and C such that $A\neq 0, C\neq B,$ but $AC=AB$.

Answer

Let $A=\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}$, $B=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$, $C=\begin{bmatrix}2 & 2\\ 2 & 2\end{bmatrix}$.

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

Exercise 2.E.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.E.18

Find 2×2 matrices A and B such that $A\neq 0$ and $B\neq 0$ but $AB=0$.

Answer

Let $A=\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}$, $B=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$.

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

Exercise 2.E.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.E.20

Find 2×2 matrices A and B such that $A\neq 0$ and $B\neq 0$ with $AB\neq BA$.

Answer

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

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

Exercise 2.E.21

Write the system

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

in the form $A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ where A is an appropriate matrix.

Answer

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

Exercise 2.E.22

Write the system

$$\begin{aligned} x_1 + 3x_2 + 2x_3 \\ 2x_3 + x_1 \\ 6x_3 \\ x_4 + 3x_2 + x_1 \end{aligned}$$

in the form $A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ where A is an appropriate matrix.

Answer

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

Exercise 2.E.23

Write the system

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

in the form $A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ where A is an appropriate matrix.

Answer

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

Exercise 2.E.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.

2.3

Exercise 2.E.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. $A=\begin{bmatrix} 1 & 3 & 1 \\ 0 & 2 & 4 \\ 1 & 0 & 5 \end{bmatrix}, B=\begin{bmatrix} 2 & 3 & 0 \\ -4 & 16 & 1 \\ 0 & 2 & 2 \end{bmatrix}$

2.4

Exercise 2.E.26

Suppose A and B are square matrices of the same size. Which of the following are necessarily true?

a. $(A-B)^2 = A^2 - 2AB + B^2$
b. $(AB)^2 = A^2B^2$
c. $(A+B)^2 = A^2 + 2AB + B^2$
d. $(A+B)^2 = A^2 + AB + BA + B^2$
e. $A^2B^2 = A(AB)B$
f. $(A+B)^3 = A^3 + 3A^2B + 3AB^2 + B^3$
g. $(A+B)(A-B) = A^2 - B^2$

Answer

a. Not necessarily true.
b. Not necessarily true.
c. Not necessarily true.
d. Necessarily true.
e. Necessarily true.
f. Not necessarily true.
g. Not necessarily true.

Exercise 2.E.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.

a. $-3A^T$
b. $3B - A^T$
c. $E^T B$
d. EE^T
e. $B^T B$
f. CA^T
g. $D^T BE$

Answer

a. $\begin{bmatrix} -3 & -9 & -3 \\ -6 & -6 & 3 \end{bmatrix}$
b. $\begin{bmatrix} 5 & -18 & 5 \\ -11 & 4 & 4 \end{bmatrix}$
c. $\begin{bmatrix} -7 & 1 & 5 \end{bmatrix}$
d. $\begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix}$
e. $\begin{bmatrix} 13 & -16 & 1 \\ -16 & 29 & -8 \\ 1 & -8 & 5 \end{bmatrix}$
f. $\begin{bmatrix} 5 & 7 & -1 \\ 5 & 15 & 5 \end{bmatrix}$
g. Not possible.

Exercise 2.E.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.E.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}.

Answer

If A is symmetric then $A = -A^T$. It follows that $a_{ii} = -a_{ii}$ and so each $a_{ii} = 0$.

Exercise 2.E.30

Prove 3 from Lemma 2.5.1. That is, show that for an $m \times n$ matrix A, an $m \times n$ matrix B, and scalars r, s, the following holds:

$$(rA + sB)^T = rA^T + sB^T$$

2.6

Exercise 2.E.31

Prove that $I_m A = A$ where A is an $m \times n$ matrix.

Answer

$(I_m A)_{ij} \equiv \sum_j \delta_{ik} A_{kj} = A_{ij}$

Exercise 2.E.32

Suppose $AB = AC$ and A is an invertible $n \times n$ matrix. Does it follow that $B = C$? Explain why or why not.

Answer

Yes $B = C$. Multiply $AB = AC$ on the left by A^{-1}.

Exercise 2.E.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.E.34

Give an example of a matrix A such that $A^2 = I$ and yet $A \neq I$ and $A \neq -I$.

Answer

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

2.7

Exercise 2.E.35

Let

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

Find A^{-1} if possible. If A^{-1} does not exist, explain why.

Answer

$$\begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{3}{7} & -\frac{1}{7} \\ \frac{1}{7} & \frac{2}{7} \end{bmatrix}$$

Exercise 2.E.36

Let

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

Find A^{-1} if possible. If A^{-1} does not exist, explain why.

Answer

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

Exercise 2.E.37

Add exercises text here.Let

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

Find A^{-1} if possible. If A^{-1} does not exist, explain why.

Answer

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

Exercise 2.E.38

Let

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

Find A^{-1} if possible. If A^{-1} does not exist, explain why.

Answer

$\begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}^{-1}$ does not exist. The of this matrix is $\begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 0 \end{bmatrix}$

Exercise 2.E.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.

Answer

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

Exercise 2.E.40

Let

https://math.libretexts.org/@go/page/170448

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

Find A^{-1} if possible. If A^{-1} does not exist, explain why.

Answer

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

? Exercise 2.E. 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.

Answer

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

? Exercise 2.E. 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.

Answer

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.

? Exercise 2.E. 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.

Answer

$$\begin{bmatrix}1&2&0&2\\1&1&2&0\\2&1&-3&2\\1&2&1&2\end{bmatrix}^{-1}=\begin{bmatrix}-1&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}\\3&\frac{1}{2}&-\frac{1}{2}&-\frac{5}{2}\\-1&0&0&1\\-2&-\frac{3}{4}&\frac{1}{4}&\frac{9}{4}\end{bmatrix}$$

Exercise 2.E.44

Using the inverse of the matrix, find the solution to the systems:

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

b. $$\begin{bmatrix}2&4\\1&1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}2\\0\end{bmatrix}$$

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.E.45

Using the inverse of the matrix, find the solution to the systems:

a. $$\begin{bmatrix}1&0&3\\2&3&4\\1&0&2\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}1\\0\\1\end{bmatrix}$$

b. $$\begin{bmatrix}1&0&3\\2&3&4\\1&0&2\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}3\\-1\\-2\end{bmatrix}$$

Now give the solution in terms of a, b, and c to the following:

$$\begin{bmatrix}1&0&3\\2&3&4\\1&0&2\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}a\\b\\c\end{bmatrix}$$

Answer

a. $$\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}1\\-\frac{2}{3}\\0\end{bmatrix}$$

b. $$\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}-12\\1\\5\end{bmatrix}$$

c. $$\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}3c-2a\\\frac{1}{3}b-\frac{2}{3}c\\a-c\end{bmatrix}$$

Exercise 2.E.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 $X=A^{-1}B$.

Answer

Multiply both sides of $AX = B$ on the left by A^{-1}.

Exercise 2.E.47

Prove that if A^{-1} exists and $AX = 0$ then $X = 0$.

Answer

Multiply on both sides on the left by A^{-1}. Thus

$$0 = A^{-1}0 = A^{-1}(AX) = \left(A^{-1}A\right)X = IX = X$$

Exercise 2.E.48: Inverse Product

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 $B = A^{-1}$.

Answer

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

Exercise 2.E.49

Show that if A is an invertible $n \times n$ matrix, then so is A^T and $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.

Answer

You need to show that $\left(A^{-1}\right)^T$ 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 $\left(A^{-1}\right)^T = \left(A^T\right)^{-1}$ and this last matrix exists.

Exercise 2.E.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.E.48.

Answer

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

Exercise 2.E.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.E. 48.

Answer

The proof of this exercise follows from the previous one.

? Exercise 2.E. 52

If A is invertible, show $\left(A^2\right)^{-1}=\left(A^{-1}\right)^2$. **Hint:** Use Problem 2.E. 48.

Answer

$A^2\left(A^{-1}\right)^2=AAA^{-1}A^{-1}=AIA^{-1}=AA^{-1}=I \quad \left(A^{-1}\right)^2A^2=A^{-1}A^{-1}AA=A^{-1}IA=A^{-1}A=I$

? Exercise 2.E. 53

If A is invertible, show $\left(A^{-1}\right)^{-1}=A$. **Hint:** Use Problem 2.E. 48.

Answer

$A^{-1}A=AA^{-1}=I$ and so by uniqueness, $\left(A^{-1}\right)^{-1}=A$.

2.8

? Exercise 2.E. 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.E. 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.E. 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.E. 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 $B=\begin{bmatrix}1&2&1\\2&-1&4\\0&5&1\end{bmatrix}$.

a. Find the elementary matrix E such that $EA=B$.
b. Find the inverse of E, E^{-1}, such that $E^{-1}B=A$.

Exercise 2.E.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 $B = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 10 & 2 \\ 2 & -1 & 4 \end{bmatrix}$.

a. Find the elementary matrix E such that $EA = B$.
b. Find the inverse of E, E^{-1}, such that $E^{-1}B = A$.

Exercise 2.E.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 $B = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 5 & 1 \\ 1 & -\frac{1}{2} & 2 \end{bmatrix}$.

a. Find the elementary matrix E such that $EA = B$.
b. Find the inverse of E, E^{-1}, such that $E^{-1}B = A$.

Exercise 2.E.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 $B = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 5 \\ 2 & -1 & 4 \end{bmatrix}$.

a. Find the elementary matrix E such that $EA = B$.
b. Find the inverse of E, E^{-1}, such that $E^{-1}B = A$.

2.10

Exercise 2.E.61

Find an LU factorization of $\begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 3 \\ 1 & 2 & 3 \end{bmatrix}$.

Answer

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

Exercise 2.E.62

Find an LU factorization of $\begin{bmatrix} 1 & 2 & 3 & 2 \\ 1 & 3 & 2 & 1 \\ 5 & 0 & 1 & 3 \end{bmatrix}$.

Answer

$$\begin{bmatrix} 1 & 2 & 3 & 2 \\ 1 & 3 & 2 & 1 \\ 5 & 0 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 5 & -10 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & -24 & -17 \end{bmatrix}$$

https://math.libretexts.org/@go/page/170448

Exercise 2.E.63

Find an LU factorization of the matrix $\begin{bmatrix} 1 & -2 & -5 & 0 \\ -2 & 5 & 11 & 3 \\ 3 & -6 & -15 & 1 \end{bmatrix}$.

Answer

$$\begin{bmatrix} 1 & -2 & -5 & 0 \\ -2 & 5 & 11 & 3 \\ 3 & -6 & -15 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & -5 & 0 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Exercise 2.E.64

Find an LU factorization of the matrix $\begin{bmatrix} 1 & -1 & -3 & -1 \\ -1 & 2 & 4 & 3 \\ 2 & -3 & -7 & -3 \end{bmatrix}$.

Answer

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

Exercise 2.E.65

Find an LU factorization of the matrix $\begin{bmatrix} 1 & -3 & -4 & -3 \\ -3 & 10 & 10 & 10 \\ 1 & -6 & 2 & -5 \end{bmatrix}$.

Answer

$$\begin{bmatrix} 1 & -3 & -4 & -3 \\ -3 & 10 & 10 & 10 \\ 1 & -6 & 2 & -5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -3 & -4 & -3 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Exercise 2.E.66

Find an LU factorization of the matrix $\begin{bmatrix} 1 & 3 & 1 & -1 \\ 3 & 10 & 8 & -1 \\ 2 & 5 & -3 & -3 \end{bmatrix}$.

Answer

$$\begin{bmatrix} 1 & 3 & 1 & -1 \\ 3 & 10 & 8 & -1 \\ 2 & 5 & -3 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 1 & -1 \\ 0 & 1 & 5 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Exercise 2.E.67

Find an LU factorization of the matrix $\begin{bmatrix} 3 & -2 & 1 \\ 9 & -8 & 6 \\ -6 & 2 & 2 \\ 3 & 2 & -7 \end{bmatrix}$.

Answer

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

? Exercise 2.E.68

Find an LU factorization of the matrix $\begin{bmatrix} -3 & -1 & 3 \\ 9 & 9 & -12 \\ 3 & 19 & -16 \\ 12 & 40 & -26 \end{bmatrix}$.

? Exercise 2.E.69

Find an LU factorization of the matrix $\begin{bmatrix} -1 & -3 & -1 \\ 1 & 3 & 0 \\ 3 & 9 & 0 \\ 4 & 12 & 16 \end{bmatrix}$.

Answer

$$\begin{bmatrix} -1 & -3 & -1 \\ 1 & 3 & 0 \\ 3 & 9 & 0 \\ 4 & 12 & 16 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ -4 & 0 & -4 & 1 \end{bmatrix} \begin{bmatrix} -1 & -3 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & -3 \\ 0 & 0 & 0 \end{bmatrix}$$

? Exercise 2.E.70

Find the LU factorization of the coefficient matrix using Dolittle's method and use it to solve the system of equations.

$$\begin{aligned} x + 2y &= 5 \\ 2x + 3y &= 6 \end{aligned}$$

Answer

An LU factorization of the coefficient matrix is

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

First solve

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

which gives $\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 5 \\ -4 \end{bmatrix}$. Then solve

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

which says that $y = 4$ and $x = -3$.

Exercise 2.E.71

Find the LU factorization of the coefficient matrix using Dolittle's method and use it to solve the system of equations.

$$\begin{aligned} x+2y+z&=1\\ y+3z&=2\\ 2x+3y&=6 \end{aligned}$$

Answer

An LU factorization of the coefficient matrix is

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

First solve

$$\begin{bmatrix}1&0&0\\0&1&0\\2&-1&1\end{bmatrix}\begin{bmatrix}u\\v\\w\end{bmatrix}=\begin{bmatrix}1\\2\\6\end{bmatrix}$$

which yields $u=1, v=2, w=6$. Next solve

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

This yields $z=6, y=-16, x=27$.

Exercise 2.E.72

Find the LU factorization of the coefficient matrix using Dolittle's method and use it to solve the system of equations.

$$\begin{aligned} x+2y+3z&=5\\ 2x+3y+z&=6\\ x-y+z&=2 \end{aligned}$$

Exercise 2.E.73

Find the LU factorization of the coefficient matrix using Dolittle's method and use it to solve the system of equations.

$$\begin{aligned} x+2y+3z&=5\\ 2x+3y+z&=6\\ 3x+5y+4z&=11 \end{aligned}$$

Answer

An LU factorization of the coefficient matrix is

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

First solve

$$\begin{bmatrix}1&0&0\\2&1&0\\3&1&1\end{bmatrix}\begin{bmatrix}u\\v\\w\end{bmatrix}=\begin{bmatrix}5\\6\\11\end{bmatrix}$$

https://math.libretexts.org/@go/page/170448

Solution is: $\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 5 \\ -4 \\ 0 \end{bmatrix}$. Next solve

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

Solution is: $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 7t-3 \\ 4-5t \\ t \end{bmatrix}, t \in \mathbb{R}$.

Exercise $2.E.74$

Is there only one LU factorization for a given matrix? **Hint:** Consider the equation

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

Look for all possible LU factorizations.

Answer

Sometimes there is more than one LU factorization as is the case in this example. The given equation clearly gives an LU factorization. However, it appears that the following equation gives another LU factorization.

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

This page titled 2.E: Exercises is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.E: Exercises** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

CHAPTER OVERVIEW

3: Determinants

3.1: Basic Techniques
3.2: Properties of Determinants
3.3: Finding Determinants using Row Operations
3.4: Applications of the Determinant
3.E: Exercises

This page titled 3: Determinants is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

3.1: Basic Techniques

Cofactors and 2 x 2 Determinants

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.

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 3.1.1: 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 3.1.1: A Two by Two Determinant

Find $\det(A)$ for the matrix $A = \begin{bmatrix} 2 & 4 \\ -1 & 6 \end{bmatrix}$.

Solution

From Definition 3.1.1,

$$\det(A) = (2)(6) - (-1)(4) = 12 + 4 = 16$$

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 3.1.2: 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 ith row and the jth 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 ith row and the jth column of A.

Hence, there is a minor associated with each entry of A. Consider the following example which demonstrates this definition.

Example 3.1.2: Finding Minors of a Matrix

Let

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

Find $minor(A)_{12}$ and $minor(A)_{23}$.

Solution

First we will find $minor(A)_{12}$. By Definition 3.1.2, 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

$$minor(A)_{12} = \det \begin{bmatrix} 4 & 2 \\ 3 & 1 \end{bmatrix}$$

Using Definition 3.1.1, we see that

$$\det \begin{bmatrix} 4 & 2 \\ 3 & 1 \end{bmatrix} = (4)(1) - (3)(2) = 4 - 6 = -2$$

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 3.1.3: The ij^{th} Cofactor of a Matrix

Suppose A is an $n \times n$ matrix. The ij^{th} **cofactor**, denoted by $\mathrm{cof}(A)_{ij}$ is defined to be

$$\mathrm{cof}(A)_{ij} = (-1)^{i+j} minor(A)_{ij}$$

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 $\mathrm{cof}(A)_{ij}$.

Example 3.1.3: Finding Cofactors of a Matrix

Consider the matrix

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

Find $\mathrm{cof}(A)_{12}$ and $\mathrm{cof}(A)_{23}$.

Solution

We will use Definition 3.1.3 to compute these cofactors.

First, we will compute $\mathrm{cof}(A)_{12}$. 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,

$$\text{cof}(A)_{12} = (-1)^{1+2} minor(A)_{12} = (-1)^{1+2}(-2) = 2$$

Hence, $\text{cof}(A)_{12} = 2$.

Similarly, we can find $\text{cof}(A)_{23}$. 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,

$$\text{cof}(A)_{23} = (-1)^{2+3} minor(A)_{23} = (-1)^{2+3}(-4) = 4$$

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 3.1.4: 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 ith row (column) is known as **expanding along the ith row (column)** as is given by

$$\det(A) = a_{i1}\text{cof}(A)_{i1} + a_{i2}\text{cof}(A)_{i2} + a_{i3}\text{cof}(A)_{i3}$$

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 3.1.4: 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 3.1.4, we take the 1 in the first column and multiply it by its cofactor,

$$1(-1)^{1+1}\begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix} = (1)(1)(-1) = -1$$

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.

$$\det(A) = \overbrace{1(-1)^{1+1}\begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix}}^{\text{cof}(A)_{11}} + \overbrace{4(-1)^{2+1}\begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix}}^{\text{cof}(A)_{21}} + \overbrace{3(-1)^{3+1}\begin{vmatrix} 2 & 3 \\ 3 & 2 \end{vmatrix}}^{\text{cof}(A)_{31}}$$

Calculating each of these, we obtain

$$\det(A) = 1(1)(-1) + 4(-1)(-4) + 3(1)(-5) = -1 + 16 + -15 = 0$$

Hence, $\det(A) = 0$.

As mentioned in Definition 3.1.4, 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.

$$\det(A) = \overbrace{4(-1)^{2+1}\begin{vmatrix}2 & 3\\ 2 & 1\end{vmatrix}}^{\text{cof}(A)_{21}} + \overbrace{3(-1)^{2+2}\begin{vmatrix}1 & 3\\ 3 & 1\end{vmatrix}}^{\text{cof}(A)_{22}} + \overbrace{2(-1)^{2+3}\begin{vmatrix}1 & 2\\ 3 & 2\end{vmatrix}}^{\text{cof}(A)_{23}}$$

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 $\det(A) = 0$.

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 3.1.1: 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 3.1.2, Definition 3.1.3 and Definition 3.1.4 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 ith row and the jth 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 3.1.5: Determinant of a Four by Four Matrix

Find $\det(A)$ where

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

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 3.1.5: 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

$$\det(A) = \sum_{j=1}^{n} a_{ij}\text{cof}(A)_{ij} = \sum_{i=1}^{n} a_{ij}\text{cof}(A)_{ij}$$

The first formula consists of expanding the determinant along the ith row and the second expands the determinant along the jth column.

In the following sections, we will explore some important properties and characteristics of the determinant.

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 3.1.6: 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.

$$\begin{bmatrix} * & * & \cdots & * \\ 0 & * & \cdots & \vdots \\ \vdots & \vdots & \ddots & * \\ 0 & \cdots & 0 & * \end{bmatrix}$$

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 3.1.2: 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 3.1.6: Determinant of a Triangular Matrix

Let

$$A = \begin{bmatrix} 1 & 2 & 3 & 77 \\ 0 & 2 & 6 & 7 \\ 0 & 0 & 3 & 33.7 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

Find $\det(A)$.

Solution

From Theorem 3.1.2, 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 3.1.2, you could use Laplace Expansion. We will expand along the first column. This gives

$$\det(A) = 1\begin{vmatrix} 2 & 6 & 7 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix} + 0(-1)^{2+1}\begin{vmatrix} 2 & 3 & 77 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix} + 0(-1)^{3+1}\begin{vmatrix} 2 & 3 & 77 \\ 2 & 6 & 7 \\ 0 & 0 & -1 \end{vmatrix} + 0(-1)^{4+1}\begin{vmatrix} 2 & 3 & 77 \\ 2 & 6 & 7 \\ 0 & 3 & 33.7 \end{vmatrix}$$

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 3.1.1 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 3.1.2 provides a much quicker method.

In the next section, we explore some important properties of determinants.

This page titled 3.1: Basic Techniques is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **3.1: Basic Techniques** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

3.2: Properties of Determinants

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 3.2.1: 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 3.2.1: 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 3.2.1: 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 3.1.1, $\det(A) = 1 \times 4 - 3 \times 2 = -2$. Notice that the rows of B are the rows of A but switched. By Theorem 3.2.1 since two rows of A have been switched, $\det(B) = -\det(A) = -(-2) = 2$. You can verify this using Definition 3.1.1.

The next theorem demonstrates the effect on the determinant of a matrix when we multiply a row by a scalar.

Theorem 3.2.2: 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 3.2.3: 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 3.2.2: 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 3.1.1, $\det(A) = -2$. We can also compute $\det(B)$ using Definition 3.1.1, and we see that $\det(B) = -10$.

Now, let's compute $\det(B)$ using Theorem 3.2.2 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 3.2.2, $\det(B) = 5 \times \det(A) = 5 \times -2 = -10$.

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 3.2.4: 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 3.2.4, 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 3.2.3: 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 3.1.1, $\det(A) = -2$. Notice that the second row of B is two times the first row of A added to the second row. By Theorem 3.2.1, $\det(B) = \det(A) = -2$. As usual, you can verify this answer using Definition 3.1.1.

Example 3.2.4: Multiple of a Row

Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. Show that $\det(A) = 0$.

Solution

Using Definition 3.1.1, 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 3.2.4 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 3.2.1 can be done with columns instead of rows. In this case, in Theorems 3.2.1, 3.2.2, and 3.2.4 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 3.2.5: 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 3.2.5: The Determinant of a Product

Compare $\det(AB)$ and $\det(A)\det(B)$ for

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

Solution

First compute AB, which is given by

$$AB = \begin{bmatrix} 1 & 2 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 4 \\ -1 & -4 \end{bmatrix}$$

and so by Definition 3.1.1

$$\det(AB) = \det \begin{bmatrix} 11 & 4 \\ -1 & -4 \end{bmatrix} = -40$$

Now

$$\det(A) = \det \begin{bmatrix} 1 & 2 \\ -3 & 2 \end{bmatrix} = 8$$

and

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

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 3.2.6: 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 3.2.6: 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 3.1.1, 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 3.2.7: 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 3.2.7: 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 3.1.1 we can find the determinant as follows:

$$\det(A) = 3\times 4 - 2\times 6 = 12 - 12 = 0$$

By Theorem 3.2.7 A is not invertible.

Now consider the matrix B. Again by Definition 3.1.1 we have

$$\det(B) = 2\times 1 - 5\times 3 = 2 - 15 = -13$$

By Theorem 3.2.7 B is invertible and the determinant of the inverse is given by

$$\begin{aligned} \det\left(A^{-1}\right) &= \frac{1}{\det(A)} \\ &= \frac{1}{-13} \\ &= -\frac{1}{13} \end{aligned}$$

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:

$$\det A = \sum_{i=1}^{n} a_{1,i}\mathrm{cof}(A)_{1,i}. \tag{3.2.1}$$

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 3.2.8:

(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 3.2.1:

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 $\operatorname{cof}(A)_{1,j}$ has a row consisting of zeros, and by our inductive assumption $\operatorname{cof}(A)_{1,j} = 0$.

On the other hand, if $j = i$ then $a_{1,j} = 0$. Therefore $a_{1,j}\operatorname{cof}(A)_{1,j} = 0$ for all j and by (3.2.1) we have

$$\det A = \sum_{j=1}^{n} a_{1,j}\operatorname{cof}(A)_{1,j} = 0$$

as each of the summands is equal to 0.

Lemma 3.2.2:

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 $\operatorname{cof}(A)_{1j} = \operatorname{cof}(B)_{1j} + \operatorname{cof}(C)_{1j}$ for $j \neq i$.

By (3.2.1) we have (using all equalities established above)

$$\begin{aligned}\det A &= \sum_{l=1}^{n} a_{1,l}\operatorname{cof}(A)_{1,l} \\ &= \sum_{l\neq i} a_{1,l}(\operatorname{cof}(B)_{1,l} + \operatorname{cof}(C)_{1,l}) + (b_{1,i} + c_{1,i})\operatorname{cof}(A)_{1,i} \\ &= \det B + \det C\end{aligned}$$

This proves that the assertion is true for all n and completes the proof.

Theorem 3.2.8:

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 $l \in \{1, \ldots, n\} \setminus \{i, j\}$. Then $A(l)$ is obtained from $B(l)$ by interchanging two of its rows (draw a picture) and by our assumption

$$\operatorname{cof}(A)_{1,l} = -\operatorname{cof}(B)_{1,l}. \tag{3.2.2}$$

Now consider $a_{1,i}\operatorname{cof}(A)_{1,l}$. 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 $a_{1i}\operatorname{cof}(A)_{1i} = -b_{1j}\operatorname{cof}(B)_{1j}$ and $a_{1j}\operatorname{cof}(A)_{1j} = -b_{1i}\operatorname{cof}(B)_{1i}$. Putting this together with (3.2.2) into (3.2.1) we see that if in the formula for $\det A$ we change the sign of each of the summands we obtain the formula for $\det B$.

$$\det A = \sum_{l=1}^{n} a_{1l}\operatorname{cof}(A)_{1l} = -\sum_{l=1}^{n} b_{1l}B_{1l} = \det B.$$

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 $\operatorname{cof}(A)_{1l} = k\operatorname{cof}(B)_{1l}$ for $l \neq i$, and for all l we have $a_{1l}\operatorname{cof}(A)_{1l} = kb_{1l}\operatorname{cof}(B)_{1l}$. By (3.2.1), we have $\det A = k \det B$.

https://math.libretexts.org/@go/page/170451

(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 3.2.2, 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 $\det C = 0$.

Theorem 3.2.9:

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 3.2.8and Theorem 3.2.8.

If C is the reduced row-echelon form of A then we can write $A = E_1 \cdot E_2 \cdot \cdots \cdot E_m \cdot C$ for some elementary matrices $E_1, \ldots, E_m$.

Now we consider two cases.

Assume first that $C = I$. Then $A = E_1 \cdot E_2 \cdot \cdots \cdot E_m$ and $AB = E_1 \cdot E_2 \cdot \cdots \cdot E_m B$. By applying the above equality m times, and then $m-1$ times, we have that

$$\begin{aligned} \det AB &= \det E_1 \det E_2 \cdot \det E_m \cdot \det B \\ &= \det(E_1 \cdot E_2 \cdot \cdots \cdot E_m) \det B \\ &= \det A \det B. \end{aligned}$$

Now assume $C \neq I$. Since it is in reduced row-echelon form, its last row consists of zeros and by (4) of Example 3.2.8the last row of CB consists of zeros. By Lemma 3.2.1we 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

$$\det AB = \det(E_1 \cdot E_2 \cdot E_m) \cdot \det(CB) = \det(E_1 \cdot E_2 \cdot \cdots \cdot E_m) 0 = 0$$

hence $\det AB = 0 = \det A \det B$.

The same 'machine' used in the previous proof will be used again.

Theorem 3.2.10:

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 3.2.8.

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 3.2.9 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 3.2.8 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 3.1.1. It is restated below.

Theorem 3.2.11:

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

$$\det A = \sum_{i=1}^{n} a_{j,i} \text{cof}(A)_{j,i}.$$

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 3.2.8 we have

$$\det A = -\det B.$$

Now we have

$$\det B = \sum_{i=1}^{n} b_{1,i} \text{cof}(B)_{1,i}.$$

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,

$$\text{cof}(B)_{1,i} = (-1)^{1+i} minor B_{1,i} = -(-1)^{2+i} minor(A)_{2,i} = -\text{cof}(A)_{2,i}$$

hence $\det B = -\sum_{i=1}^{n} a_{2,i} \text{cof}(A)_{2,i}$, and therefore $\det A = -\det B = \sum_{i=1}^{n} a_{2,i} \text{cof}(A)_{2,i}$ as desired.

The case when $j > 2$ is very similar; we still have $minor(B)_{1,i} = minor(A)_{j,i}$ but checking that $\det B = -\sum_{i=1}^{n} a_{j,i} \text{cof}(A)_{j,i}$ 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 3.2.10, 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.

This page titled 3.2: Properties of Determinants is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **3.2: Properties of Determinants** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

3.3: Finding Determinants using Row Operations

Theorems 3.2.1, 3.2.2 and 3.2.4 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 3.3.1: Finding a Determinant

Find the determinant of the matrix

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

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

$$B = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -9 & -13 & -17 \\ 0 & -3 & -8 & -13 \\ 0 & -2 & -10 & -3 \end{bmatrix}$$

Notice that the only row operation we have done so far is adding a multiple of a row to another row. Therefore, by Theorem 3.2.4, $\det(B) = \det(A)$.

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 3.2.4 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 3.2.4. 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 $\det(A) = \left(\frac{1}{3}\right)(1485) = 495$.

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 3.3.1, 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 3.3.2: 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 3.2.4, $\det(A) = \det(B)$.

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 3.2.4 $\det(A) = \det(C)$.

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 3.2.4 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.

This page titled 3.3: Finding Determinants using Row Operations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **3.3: Finding Determinants using Row Operations** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

3.4: 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.

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 2.6.1. 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 3.4.1: The Cofactor Matrix

Let $A = [a_{ij}]$ be an $n \times n$ matrix. Then the **cofactor matrix of A**, denoted $\mathrm{cof}(A)$, is defined by $\mathrm{cof}(A) = \left[\mathrm{cof}(A)_{ij}\right]$ where $\mathrm{cof}(A)_{ij}$ is the ij^{th} cofactor of A.

Note that $\mathrm{cof}(A)_{ij}$ 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 $adj(A)$.

In the specific case where A is a 2×2 matrix given by

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

then $adj(A)$ is given by

$$adj(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

In general, $adj(A)$ 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 3.4.1: 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 3.4.1: 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 3.4.1.

Solution

According to Theorem 3.4.1,

$$A^{-1} = \frac{1}{\det(A)} adj(A)$$

First we will find the determinant of this matrix. Using Theorems 3.2.1, 3.2.2, and 3.2.4, 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 3.2.4, $\det(A) = \det(B)$. By Theorem 3.1.2, $\det(B) = 1 \times -6 \times -2 = 12$. Hence, $\det(A) = 12$.

Now, we need to find $adj(A)$. To do so, first we will find the cofactor matrix of A. This is given by

$$\text{cof}(A) = \begin{bmatrix} -2 & -2 & 6 \\ 4 & -2 & 0 \\ 2 & 8 & -6 \end{bmatrix}$$

Here, the ij^{th} entry is the ij^{th} cofactor of the original matrix A which you can verify. Therefore, from Theorem 3.4.1, 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.

We will look at another example of how to use this formula to find A^{-1}.

Example 3.4.2: Find the Inverse From a Formula

Find the inverse of the matrix

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

using the formula given in Theorem 3.4.1.

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

$$A^{-1} = \frac{1}{(1/6)}\, adj(A) = 6\, adj(A)$$

We continue to calculate as follows. Here we show the 2×2 determinants needed to find the cofactors.

$$A^{-1} = 6 \begin{bmatrix} \begin{vmatrix} \frac{1}{3} & -\frac{1}{2} \\ \frac{2}{3} & -\frac{1}{2} \end{vmatrix} & -\begin{vmatrix} -\frac{1}{6} & -\frac{1}{2} \\ -\frac{5}{6} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} -\frac{1}{6} & \frac{1}{3} \\ -\frac{5}{6} & \frac{2}{3} \end{vmatrix} \\ -\begin{vmatrix} 0 & \frac{1}{2} \\ \frac{2}{3} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{5}{6} & -\frac{1}{2} \end{vmatrix} & -\begin{vmatrix} \frac{1}{2} & 0 \\ -\frac{5}{6} & \frac{2}{3} \end{vmatrix} \\ \begin{vmatrix} 0 & \frac{1}{2} \\ \frac{1}{3} & -\frac{1}{2} \end{vmatrix} & -\begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{6} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} \frac{1}{2} & 0 \\ -\frac{1}{6} & \frac{1}{3} \end{vmatrix} \end{bmatrix}^T$$

Expanding all the 2×2 determinants, this yields

$$A^{-1} = 6 \begin{bmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{6} & -\frac{1}{3} \\ -\frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}^T = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -2 & 1 \end{bmatrix}$$

Again, you can always check your work by multiplying $A^{-1}A$ and AA^{-1} and ensuring these products equal I.

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

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 3.4.1.

Theorem 3.4.1: The Inverse and the Determinant

Proof

(of Theorem 3.4.1) Recall that the (i, j)-entry of $adj(A)$ is equal to $\text{cof}(A)_{ji}$. Thus the (i, j)-entry of $B = A \cdot adj(A)$ is :

$$B_{ij} = \sum_{k=1}^{n} a_{ik} adj(A)_{kj} = \sum_{k=1}^{n} a_{ik} \text{cof}(A)_{jk}$$

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:

$$A\, adj(A) = \det(A) I$$

Similarly we can verify that:

$$adj(A)\, A = \det(A) I$$

And this proves the first part of the theorem.

Further if A is invertible, then by Theorem 3.2.5 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:

$$A^{-1} = \frac{1}{\det(A)} adj(A)$$

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 3.4.3: Inverse for Non-Constant Matrix

Suppose

$$A(t) = \begin{bmatrix} e^t & 0 & 0 \\ 0 & \cos t & \sin t \\ 0 & -\sin t & \cos t \end{bmatrix}$$

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

$$\frac{1}{e^t}\begin{bmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{bmatrix}^T = \begin{bmatrix} e^{-t} & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{bmatrix}$$

Cramer's Rule

Another context in which the formula given in Theorem 3.4.1 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

$$X = A^{-1}B = \frac{1}{\det(A)} adj(A)\, B$$

Let x_i be the ith entry of X and b_j be the jth entry of B. Then this equation becomes

$$x_i = \sum_{j=1}^{n} [a_{ij}]^{-1} b_j = \sum_{j=1}^{n} \frac{1}{\det(A)} adj(A)_{ij} b_j$$

where $adj(A)_{ij}$ is the ij^{th} entry of $adj(A)$.

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 ith column of A is replaced with the column vector $[b_1 \cdots, b_n]^T$. 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 3.4.1: 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 ith 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 3.4.4: 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 3.4.1 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.

$$z = \frac{\det(A_3)}{\det(A)} = \frac{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 2 \\ 2 & -3 & 3 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{vmatrix}} = \frac{11}{14}$$

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.

Using Cramer's Rule

Example 3.4.5: Use Cramer's Rule for Non-Constant Matrix

Solve for z if

https://math.libretexts.org/@go/page/170453

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ t \\ 0.05int^2 \end{bmatrix}$$

Solution

We are asked to find the value of z in the solution. We will solve using Cramer's rule. Thus

$$z = .05in \frac{\begin{vmatrix} 1 & 0 & 1 \\ 0 & e^t \cos t & t \\ 0 & -e^t \sin t & t^2 \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{vmatrix}} = t\left((\cos t)\, t + \sin t\right) e^{-t}$$

Polynomial Interpolation

In studying a set of data that relates variables x and y, it may be the case that we can use a polynomial to "fit" to the data. If such a polynomial can be established, it can be used to estimate values of x and y which have not been provided.

Consider the following example.

Example 3.4.6: Polynomial Interpolation

Given data points $(1,4),(2,9),(3,12)$ find an interpolating polynomial $p(x)$ of degree at most 2 and then estimate the value corresponding to $x = \frac{1}{2}$.

Solution

We want to find a polynomial given by

$$p(x) = r_0 + r_1 x_1 + r_2 x_2^2$$

such that $p(1) = 4, p(2) = 9$ and $p(3) = 12$. To find this polynomial, substitute the known values in for x and solve for r_0, r_1, and r_2.

$$\begin{aligned} p(1) &= r_0 + r_1 + r_2 = 4 \\ p(2) &= r_0 + 2r_1 + 4r_2 = 9 \\ p(3) &= r_0 + 3r_1 + 9r_2 = 12 \end{aligned}$$

Writing the augmented matrix, we have

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 1 & 2 & 4 & 9 \\ 1 & 3 & 9 & 12 \end{array}\right]$$

After row operations, the resulting matrix is

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

Therefore the solution to the system is $r_0 = -3, r_1 = 8, r_2 = -1$ and the required interpolating polynomial is

$$p(x) = -3 + 8x - x^2$$

To estimate the value for $x = \frac{1}{2}$, we calculate $p(\frac{1}{2})$:

$$\begin{aligned} p(\frac{1}{2}) &= -3+8(\frac{1}{2})-(\frac{1}{2})^2 \\ &= -3+4-\frac{1}{4} \\ &= \frac{3}{4} \end{aligned}$$

This procedure can be used for any number of data points, and any degree of polynomial. The steps are outlined below.

Procedure 3.4.2: Finding an Interpolation Polynomial

Suppose that values of x and corresponding values of y are given, such that the actual relationship between x and y is unknown. Then, values of y can be estimated using an **interpolating polynomial $p(x)$**. If given $x_1, \ldots, x_n$ and the corresponding $y_1, \ldots, y_n$, the procedure to find $p(x)$ is as follows:

1. The desired polynomial $p(x)$ is given by
$$p(x) = r_0 + r_1 x + r_2 x^2 + \ldots + r_{n-1} x^{n-1}$$
2. $p(x_i) = y_i$ for all $i = 1, 2, \ldots, n$ so that
$$\begin{aligned} r_0 + r_1 x_1 + r_2 x_1^2 + \ldots + r_{n-1} x_1^{n-1} &= y_1 \\ r_0 + r_1 x_2 + r_2 x_2^2 + \ldots + r_{n-1} x_2^{n-1} &= y_2 \\ &\vdots \\ r_0 + r_1 x_n + r_2 x_n^2 + \ldots + r_{n-1} x_n^{n-1} &= y_n \end{aligned}$$
3. Set up the augmented matrix of this system of equations
$$\left[\begin{array}{ccccc|c} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} & y_1 \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} & y_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} & y_n \end{array}\right]$$
4. Solving this system will result in a unique solution $r_0, r_1, \cdots, r_{n-1}$. Use these values to construct $p(x)$, and estimate the value of $p(a)$ for any $x = a$.

This procedure motivates the following theorem.

Theorem 3.4.2: Polynomial Interpolation

Given n data points $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ with the x_i distinct, there is a unique polynomial $p(x) = r_0 + r_1 x + r_2 x^2 + \cdots + r_{n-1} x^{n-1}$ such that $p(x_i) = y_i$ for $i = 1, 2, \cdots, n$. The resulting polynomial $p(x)$ is called the **interpolating polynomial** for the data points.

We conclude this section with another example.

Example 3.4.7: Polynomial Interpolation

Consider the data points $(0, 1), (1, 2), (3, 22), (5, 66)$. Find an interpolating polynomial $p(x)$ of degree at most three, and estimate the value of $p(2)$.

Solution

The desired polynomial $p(x)$ is given by:

$$p(x) = r_0 + r_1 x + r_2 x^2 + r_3 x^3$$

Using the given points, the system of equations is

$$\begin{aligned} p(0) &= r_0 = 1 \\ p(1) &= r_0 + r_1 + r_2 + r_3 = 2 \\ p(3) &= r_0 + 3r_1 + 9r_2 + 27r_3 = 22 \\ p(5) &= r_0 + 5r_1 + 25r_2 + 125r_3 = 66 \end{aligned}$$

The augmented matrix is given by:

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 2 \\ 1 & 3 & 9 & 27 & 22 \\ 1 & 5 & 25 & 125 & 66 \end{array}\right]$$

The resulting matrix is

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

Therefore, $r_0 = 1, r_1 = -2, r_2 = 3, r_3 = 0$ and $p(x) = 1 - 2x + 3x^2$. To estimate the value of $p(2)$, we compute $p(2) = 1 - 2(2) + 3(2^2) = 1 - 4 + 12 = 9$.

This page titled 3.4: Applications of the Determinant is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **3.4: Applications of the Determinant** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

3.E: Exercises

Exercise 3.E.1

Find the determinants of the following matrices.

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

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

c. $\begin{bmatrix} 4 & 3 \\ 6 & 2 \end{bmatrix}$

Exercise 3.E.2

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

a. $minor(A)_{11}$

b. $minor(A)_{21}$

c. $minor(A)_{32}$

d. $cof(A)_{11}$

e. $cof(A)_{21}$

f. $cof(A)_{32}$

Exercise 3.E.3

Find the determinants of the following matrices.

a. $\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 2 \\ 0 & 9 & 8 \end{bmatrix}$

b. $\begin{bmatrix} 4 & 3 & 2 \\ 1 & 7 & 8 \\ 3 & -9 & 3 \end{bmatrix}$

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

Answer

a. The answer is 31.

b. The answer is 375.

c. The answer is -2.

Exercise 3.E.4

Find the following determinant by expanding along the first row and second column.

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix}$$

Answer

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix} = 6$$

? Exercise $3.E.5$

Find the following determinant by expanding along the first column and third row.

$$\begin{vmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{vmatrix}$$

Answer

$$\begin{vmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{vmatrix} = 2$$

? Exercise $3.E.6$

Find the following determinant by expanding along the second row and first column.

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix}$$

Answer

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix} = 6$$

? Exercise $3.E.7$

Compute the determinant by cofactor expansion. Pick the easiest row or column to use.

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

Answer

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

? Exercise $3.E.8$

Find the determinant of the following matrices.

a. $A = \begin{bmatrix} 1 & -34 \\ 0 & 2 \end{bmatrix}$

b. $A = \begin{bmatrix} 4 & 3 & 14 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

c. $A = \begin{bmatrix} 2 & 3 & 15 & 0 \\ 0 & 4 & 1 & 7 \\ 0 & 0 & -3 & 5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

Exercise 3.E.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.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \to \cdots \to \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

Answer

It does not change the determinant. This was just taking the transpose.

Exercise 3.E.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.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \to \cdots \to \begin{bmatrix} c & d \\ a & b \end{bmatrix}$$

Answer

In this case two rows were switched and so the resulting determinant is -1 times the first

Exercise 3.E.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.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \to \cdots \to \begin{bmatrix} a & b \\ a+c & b+d \end{bmatrix}$$

Answer

The determinant is unchanged. It was just the first row added to the second.

Exercise 3.E.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.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \to \cdots \to \begin{bmatrix} a & b \\ 2c & 2d \end{bmatrix}$$

Answer

The second row was multiplied by 2 so the determinant of the result is 2 times the original determinant.

Exercise 3.E.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.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \to \cdots \to \begin{bmatrix} b & a \\ d & c \end{bmatrix}$$

Answer

In this case the two columns were switched so the determinant of the second is -1 times the determinant of the first.

Exercise 3.E.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 $\det(A) = 0$.

Answer

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.

Exercise 3.E.15

Show $\det(aA) = a^n \det(A)$ for an $n \times n$ matrix A and scalar a.

Answer

$\det(aA) = \det(aIA) = \det(aI)\det(A) = a^n \det(A)$. The matrix which has a down the main diagonal has determinant equal to a^n.

Exercise 3.E.16

Construct 2×2 matrices A and B to show that the $\det A \det B = \det(AB)$.

Answer

$$\det\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} -1 & 2 \\ -5 & 6 \end{bmatrix}\right) = -8$$

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

Exercise 3.E.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.

Answer

This is not true at all. Consider $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$.

Exercise 3.E.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 $\det(A)$?

Answer

It must be 0 because $0 = \det(0) = \det(A^k) = (\det(A))^k$.

Exercise 3.E.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?

Answer

You would need $\det(AA^T) = \det(A)\det(A^T) = \det(A)^2 = 1$ and so $\det(A) = 1$, or -1.

Exercise 3.E.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 $\det(A) = \det(B)$.

Answer

$\det(A) = \det(S^{-1}BS) = \det(S^{-1})\det(B)\det(S) = \det(B)\det(S^{-1}S) = \det(B)$.

Exercise 3.E.21

Tell whether each statement is true or false. If true, provide a proof. If false, provide a counter example.

a. If A is a 3×3 matrix with a zero determinant, then one column must be a multiple of some other column.
b. If any two columns of a square matrix are equal, then the determinant of the matrix equals zero.
c. For two $n \times n$ matrices A and B, $\det(A+B) = \det(A) + \det(B)$.
d. For an $n \times n$ matrix A, $\det(3A) = 3\det(A)$
e. If A^{-1} exists then $\det(A^{-1}) = \det(A)^{-1}$.
f. If B is obtained by multiplying a single row of A by 4 then $\det(B) = 4\det(A)$.
g. For A an $n \times n$ matrix, $\det(-A) = (-1)^n \det(A)$.
h. If A is a real $n \times n$ matrix, then $\det(A^T A) \geq 0$.
i. If $A^k = 0$ for some positive integer k, then $\det(A) = 0$.
j. If $AX = 0$ for some $X \neq 0$, then $\det(A) = 0$.

Answer

a. False. Consider $\begin{bmatrix} 1 & 1 & 2 \\ -1 & 5 & 4 \\ 0 & 3 & 3 \end{bmatrix}$
b. True.
c. False.
d. False.
e. True.
f. False.
g. True.
h. True.
i. True.
j. True.

Exercise 3.E.22

Find the determinant using row operations to first simplify.

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

Answer

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

Exercise 3.E.23

Find the determinant using row operations to first simplify.

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

Answer

$$\begin{vmatrix} 2 & 1 & 3 \\ 2 & 4 & 2 \\ 1 & 4 & -5 \end{vmatrix} = -32$$

Exercise 3.E.24

Find the determinant using row operations to first simplify.

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

Answer

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.

$$\begin{vmatrix} 1 & 2 & 1 & 2 \\ 3 & 1 & -2 & 3 \\ -1 & 0 & 3 & 1 \\ 2 & 3 & 2 & -2 \end{vmatrix} = 63$$

https://math.libretexts.org/@go/page/170454

Exercise $3.E.25$

Find the determinant using row operations to first simplify.

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

Answer

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

$$\begin{vmatrix} 1 & 4 & 1 & 2 \\ 3 & 2 & -2 & 3 \\ -1 & 0 & 3 & 3 \\ 2 & 1 & 2 & -2 \end{vmatrix} = 211$$

Exercise $3.E.26$

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.

Answer

$\det \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 3 & 1 & 0 \end{bmatrix} = -13$ and so it has an inverse. This inverse is

$$\frac{1}{-13}\begin{bmatrix} \begin{vmatrix} 2 & 1 \\ 1 & 0 \end{vmatrix} & -\begin{vmatrix} 0 & 1 \\ 3 & 0 \end{vmatrix} & \begin{vmatrix} 0 & 2 \\ 3 & 1 \end{vmatrix} \\ -\begin{vmatrix} 2 & 3 \\ 1 & 0 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 3 & 0 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} \\ \begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 0 & 1 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 0 & 2 \end{vmatrix} \end{bmatrix}^T = \frac{1}{-13}\begin{bmatrix} -1 & 3 & -6 \\ 3 & -9 & 5 \\ -4 & -1 & 2 \end{bmatrix}^T$$

$$= \begin{bmatrix} \frac{1}{13} & -\frac{3}{13} & \frac{4}{13} \\ -\frac{3}{13} & \frac{9}{13} & \frac{1}{13} \\ \frac{6}{13} & -\frac{5}{13} & -\frac{2}{13} \end{bmatrix}$$

https://math.libretexts.org/@go/page/170454

Exercise 3.E.27

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.

Answer

$\det \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 1 \\ 3 & 1 & 1 \end{bmatrix} = 7$ so it has an inverse. This inverse is $\frac{1}{7} \begin{bmatrix} 1 & 3 & -6 \\ -2 & 1 & 5 \\ 2 & -1 & 2 \end{bmatrix}^T = \begin{bmatrix} \frac{1}{7} & -\frac{2}{7} & \frac{2}{7} \\ \frac{3}{7} & \frac{1}{7} & -\frac{1}{7} \\ -\frac{6}{7} & \frac{5}{7} & \frac{2}{7} \end{bmatrix}$

Exercise 3.E.28

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.

Answer

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

so it has an inverse which is

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

Exercise 3.E.29

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.E.30

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.

Answer

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

and so it has an inverse. The inverse turns out to equal

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

Exercise $3.E.31$

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.

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

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

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

Answer

a. $\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} = 1$

b. $\begin{vmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 4 & 1 & 1 \end{vmatrix} = -15$

c. $\begin{vmatrix} 1 & 2 & 1 \\ 2 & 3 & 0 \\ 0 & 1 & 2 \end{vmatrix} = 0$

Exercise $3.E.32$

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.

Answer

https://math.libretexts.org/@go/page/170454

No. It has a nonzero determinant for all t

Exercise 3.E.33

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.

Answer

$$\det \begin{bmatrix} 1 & t & t^2 \\ 0 & 1 & 2t \\ t & 0 & 2 \end{bmatrix} = t^3 + 2$$

and so it has no inverse when $t = -\sqrt[3]{2}$

Exercise 3.E.34

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.

Answer

$$\det \begin{bmatrix} e^t & \cosh t & \sinh t \\ e^t & \sinh t & \cosh t \\ e^t & \cosh t & \sinh t \end{bmatrix} = 0$$

and so this matrix fails to have a nonzero determinant at any value of t.

Exercise 3.E.35

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.

Answer

$$\det \begin{bmatrix} e^t & e^{-t}\cos t & e^{-t}\sin t \\ e^t & -e^{-t}\cos t - e^{-t}\sin t & -e^{-t}\sin t + e^{-t}\cos t \\ e^t & 2e^{-t}\sin t & -2e^{-t}\cos t \end{bmatrix} = 5e^{-t} \neq 0$$

and so this matrix is always invertible.

Exercise 3.E.36

Show that if $\det(A) \neq 0$ for A an $n \times n$ matrix, it follows that if $AX = 0$, then $X = 0$.

Answer

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 $X = 0$.

Exercise 3.E.37

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 Exercise 3.E.36.

Answer

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 $BA = I$.

Exercise 3.E.38

Use the formula for the inverse in terms of the cofactor matrix to find the inverse of the matrix

$$A = \begin{bmatrix} e^t & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & e^t \cos t - e^t \sin t & e^t \cos t + e^t \sin t \end{bmatrix}$$

Answer

$$\det \begin{bmatrix} e^t & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & e^t \cos t - e^t \sin t & e^t \cos t + e^t \sin t \end{bmatrix} = e^{3t}.$$

Hence the inverse is

$$e^{-3t} \begin{bmatrix} e^{2t} & 0 & 0 \\ 0 & e^{2t} \cos t + e^{2t} \sin t & -(e^{2t} \cos t - e^{2t} \sin)t \\ 0 & -e^{2t} \sin t & e^{2t} \cos(t) \end{bmatrix}^T$$
$$= \begin{bmatrix} e^{-t} & 0 & 0 \\ 0 & e^{-t}(\cos t + \sin t) & -(\sin t)e^{-t} \\ 0 & -e^{-t}(\cos t - \sin t) & (\cos t)e^{-t} \end{bmatrix}$$

Exercise 3.E.39

Find the inverse, if it exists, of the matrix

$$A = \begin{bmatrix} e^t & \cos t & \sin t \\ e^t & -\sin t & \cos t \\ e^t & -\cos t & -\sin t \end{bmatrix}$$

Answer

$$\begin{bmatrix} e^t & \cos t & \sin t \\ e^t & -\sin t & \cos t \\ e^t & -\cos t & -\sin t \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} \frac{1}{2}e^{-t} & 0 & \frac{1}{2}e^{-t} \\ \frac{1}{2}\cos t + \frac{1}{2}\sin t & -\sin t & \frac{1}{2}\sin t - \frac{1}{2}\cos t \\ \frac{1}{2}\sin t - \frac{1}{2}\cos t & \cos t & -\frac{1}{2}\cos t - \frac{1}{2}\sin t \end{bmatrix}$$

? Exercise 3.E.40

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?

Answer

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

? Exercise 3.E.41

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.

Answer

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.

? Exercise 3.E.42

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.

Answer

False.

? Exercise 3.E.43

Use Cramer's rule to find the solution to

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

Answer

Solution is: $[x = 1, y = 0]$

? Exercise 3.E.44

Use Cramer's rule to find the solution to

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

Answer

Solution is: $[x=1, y=0, z=0]$. For example,

$$y = \frac{\begin{vmatrix} 1 & 1 & 1 \\ 2 & 2 & -1 \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 2 & -1 & -1 \\ 1 & 0 & 1 \end{vmatrix}} = 0$$

This page titled 3.E: Exercises is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

4: Linear Transformations

4.1: Linear Transformations
4.2: The Matrix of a Linear Transformation I
4.3: Properties of Linear Transformations. Invertible Linear Transformations.
4.4: Special Linear Transformations and Applications to Computer Graphics
4.5: One-to-One and Onto Transformations
4.6: Isomorphisms
4.7: The Kernel and Image of A Linear Map
4.8: The Matrix of a Linear Transformation II
4.9: The General Solution of a Linear System
4.E: Exercises

This page titled 4: Linear Transformations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

4.1: 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

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

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 4.1.1: A Function Which Transforms Vectors

Consider the matrix $A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{bmatrix}$. 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 $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ in $\mathbb{R}^3$, multiply on the left by the given matrix to obtain the new vector. This product looks like

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

The resulting product is a 2×1 vector which is determined by the choice of x and y. Here are some numerical examples.

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

Here, the vector $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ in $\mathbb{R}^3$ was transformed by the matrix into the vector $\begin{bmatrix} 5 \\ 4 \end{bmatrix}$ in $\mathbb{R}^2$.

Here is another example:

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 10 \\ 5 \\ -3 \end{bmatrix} = \begin{bmatrix} 20 \\ 25 \end{bmatrix}$$

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 4.1.1: 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 4.1.2: Linear Transformation

Let T be a transformation defined by $T:\mathbb{R}^3\to\mathbb{R}^2$ is defined by

$$T\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}x+y\\x-z\end{bmatrix}\text{ for all }\begin{bmatrix}x\\y\\z\end{bmatrix}\in\mathbb{R}^3$$

Show that T is a linear transformation.

Solution

By Definition 4.1.1 we need to show that $T(k\vec{x}_1+p\vec{x}_2)=kT(\vec{x}_1)+pT(\vec{x}_2)$ for all scalars k,p and vectors $\vec{x}_1,\vec{x}_2$. Let

$$\vec{x}_1=\begin{bmatrix}x_1\\y_1\\z_1\end{bmatrix},\vec{x}_2=\begin{bmatrix}x_2\\y_2\\z_2\end{bmatrix}$$

Then

$$\begin{aligned}
T\left(k\vec{x}_1+p\vec{x}_2\right) &= T\left(k\begin{bmatrix}x_1\\y_1\\z_1\end{bmatrix}+p\begin{bmatrix}x_2\\y_2\\z_2\end{bmatrix}\right)\\
&= T\left(\begin{bmatrix}kx_1\\ky_1\\kz_1\end{bmatrix}+\begin{bmatrix}px_2\\py_2\\pz_2\end{bmatrix}\right)\\
&= T\left(\begin{bmatrix}kx_1+px_2\\ky_1+py_2\\kz_1+pz_2\end{bmatrix}\right)\\
&= \begin{bmatrix}(kx_1+px_2)+(ky_1+py_2)\\(kx_1+px_2)-(kz_1+pz_2)\end{bmatrix}\\
&= \begin{bmatrix}(kx_1+ky_1)+(px_2+py_2)\\(kx_1-kz_1)+(px_2-pz_2)\end{bmatrix}\\
&= \begin{bmatrix}kx_1+ky_1\\kx_1-kz_1\end{bmatrix}+\begin{bmatrix}px_2+py_2\\px_2-pz_2\end{bmatrix}\\
&= k\begin{bmatrix}x_1+y_1\\x_1-z_1\end{bmatrix}+p\begin{bmatrix}x_2+y_2\\x_2-z_2\end{bmatrix}\\
&= kT(\vec{x}_1)+pT(\vec{x}_2)
\end{aligned}$$

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 4.1.2.

We began this section by discussing matrix transformations, where multiplication by a matrix transforms vectors. These matrix transformations are in fact linear transformations.

Theorem 4.1.1: 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.

This page titled 4.1: Linear Transformations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **5.1: Linear Transformations** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.2: The Matrix of a Linear Transformation I

Outcomes

A. Find the matrix of a linear transformation with respect to the standard basis.
B. Determine the action of a linear transformation 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

$$T\left(\vec{x}\right) = A\vec{x} \tag{4.2.1}$$

for all $\vec{x} \in \mathbb{R}^n$.

Theorem 4.2.1: 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 (4.2.1). Note that

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = \sum_{i=1}^{n} x_i \vec{e}_i$$

where $\vec{e}_i$ is the ith column of I_n, that is the $n \times 1$ vector which has zeros in every slot but the ith and a 1 in this slot.

Then since T is linear,

$$\begin{aligned} T\left(\vec{x}\right) &= \sum_{i=1}^{n} x_i T\left(\vec{e}_i\right) \\ &= \begin{bmatrix} | & & | \\ T\left(\vec{e}_1\right) & \cdots & T\left(\vec{e}_n\right) \\ | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ &= A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \end{aligned}$$

The desired matrix is obtained from constructing the ith column as $T\left(\vec{e}_i\right)$. Recall that the set $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$ is called the standard basis of $\mathbb{R}^n$. Therefore the matrix of T is found by applying T to the standard basis. We state this formally as the following theorem.

Theorem 4.2.2: Matrix of a Linear Transformation

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear transformation. Then the matrix A satisfying $T\left(\vec{x}\right) = A\vec{x}$ is given by

$$A = \begin{bmatrix} | & & | \\ T\left(\vec{e}_1\right) & \cdots & T\left(\vec{e}_n\right) \\ | & & | \end{bmatrix}$$

where $\vec{e}_i$ is the ith column of I_n, and then $T\left(\vec{e}_i\right)$ is the ith column of A.

The following Corollary is an essential result.

Corollary 4.2.1: Matrix and Linear Transformation

A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if it is a matrix transformation.

Consider the following example.

Example 4.2.1: The Matrix of a Linear Transformation

Suppose T is a linear transformation, $T : \mathbb{R}^3 \to \mathbb{R}^2$ where

$$T\begin{bmatrix}1\\0\\0\end{bmatrix} = \begin{bmatrix}1\\2\end{bmatrix}, T\begin{bmatrix}0\\1\\0\end{bmatrix} = \begin{bmatrix}9\\-3\end{bmatrix}, T\begin{bmatrix}0\\0\\1\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}$$

Find the matrix A of T such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x}$.

Solution

By Theorem 4.2.2 we construct A as follows:

$$A = \begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & | \end{bmatrix}$$

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,

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

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 4.2.2: The Matrix of Linear Transformation: Inconveniently Defined

Suppose T is a linear transformation, $T : \mathbb{R}^2 \to \mathbb{R}^2$ and

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

Find the matrix A of T such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x}$.

Solution

By Theorem 4.2.2 to find this matrix, we need to determine the action of T on $\vec{e}_1$ and $\vec{e}_2$. In Example 9.9.2, 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

$$\begin{bmatrix}1\\0\end{bmatrix} = x\begin{bmatrix}1\\1\end{bmatrix} + y\begin{bmatrix}0\\-1\end{bmatrix} \tag{4.2.2}$$

Then, since T is linear,

$$T\begin{bmatrix}1\\0\end{bmatrix} = xT\begin{bmatrix}1\\1\end{bmatrix} + yT\begin{bmatrix}0\\-1\end{bmatrix}$$

Substituting in values, this sum becomes

$$T\begin{bmatrix}1\\0\end{bmatrix}=x\begin{bmatrix}1\\2\end{bmatrix}+y\begin{bmatrix}3\\2\end{bmatrix} \tag{4.2.3}$$

Therefore, if we know the values of x and y which satisfy (4.2.2), we can substitute these into equation (4.2.3). 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 (4.2.2), which can be done by solving the system

$$\begin{aligned}x&=1\\x-y&=0\end{aligned}$$

We see that $x=1$ and $y=1$ is the solution to this system. Substituting these values into equation (4.2.3), we have

$$T\begin{bmatrix}1\\0\end{bmatrix}=1\begin{bmatrix}1\\2\end{bmatrix}+1\begin{bmatrix}3\\2\end{bmatrix}=\begin{bmatrix}1\\2\end{bmatrix}+\begin{bmatrix}3\\2\end{bmatrix}=\begin{bmatrix}4\\4\end{bmatrix}$$

Therefore $\begin{bmatrix}4\\4\end{bmatrix}$ 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

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

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 4.2.1: 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 $\begin{bmatrix}\vec{a}_1&\cdots&\vec{a}_n\end{bmatrix}^{-1}$ exists, and

$$T(\vec{a}_i)=\vec{b}_i$$

Then the matrix of T must be of the form

$$\begin{bmatrix}\vec{b}_1&\cdots&\vec{b}_n\end{bmatrix}\begin{bmatrix}\vec{a}_1&\cdots&\vec{a}_n\end{bmatrix}^{-1}$$

We will illustrate this procedure in the following example. You may also find it useful to work through Example 4.2.2 using this procedure.

Example 4.2.3: 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 4.2.1, $A=\begin{bmatrix}1&0&1\\3&1&1\\1&1&0\end{bmatrix}^{-1}$ and $B=\begin{bmatrix}0&2&0\\1&1&0\\1&3&1\end{bmatrix}$

Then, Procedure 4.2.1 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:

$$T(\vec{x}) = T(A\vec{y}) = T\left(\sum_i \vec{y}_i \vec{a}_i\right) = \sum \vec{y}_i T(\vec{a}_i) \sum \vec{y}_i \vec{b}_i = B\vec{y} = BA^{-1}\vec{x} = C\vec{x}$$

Recall the dot product discussed earlier. Consider the map $\vec{v} \mapsto \text{proj}_{\vec{u}}(\vec{v})$ 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.

$$\begin{aligned} \text{proj}_{\vec{u}}(k\vec{v} + p\vec{w}) &= \left(\frac{(k\vec{v} + p\vec{w}) \bullet \vec{u}}{\vec{u} \bullet \vec{u}}\right)\vec{u} \\ &= k\left(\frac{\vec{v} \bullet \vec{u}}{\vec{u} \bullet \vec{u}}\right)\vec{u} + p\left(0.05in\frac{\vec{w} \bullet \vec{u}}{\vec{u} \bullet \vec{u}}\right)\vec{u} \\ &= k\,\text{proj}_{\vec{u}}(\vec{v}) + p\,\text{proj}_{\vec{u}}(\vec{w}) \end{aligned}$$

Consider the following example.

Example 4.2.4: 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

$$T(\vec{v}) = \text{proj}_{\vec{u}}(\vec{v})$$

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 $T(\vec{v}) = \text{proj}_{\vec{u}}(\vec{v})$ is linear. Therefore by Theorem 4.2.1, 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 $T(\vec{e}_i) = \text{proj}_{\vec{u}}(\vec{e}_i)$ gives the ith column of the desired matrix. Therefore, we need to find

$$\text{proj}_{\vec{u}}(\vec{e}_i) = \left(\frac{\vec{e}_i \bullet \vec{u}}{\vec{u} \bullet \vec{u}}\right)\vec{u}$$

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

$$\frac{1}{14}\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

This page titled 4.2: The Matrix of a Linear Transformation I is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **5.2: The Matrix of a Linear Transformation I** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.3: Properties of Linear Transformations. Invertible 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 4.3.1: Properties of Linear Transformations

Properties of Linear Transformationsproperties 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:

$$\text{Let } \vec{x}_1, \ldots, \vec{x}_k \in \mathbb{R}^n \text{ and } a_1, \ldots, a_k \in \mathbb{R}.$$

$$\text{Then if } \vec{y} = a_1\vec{x}_1 + a_2\vec{x}_2 + \ldots + a_k\vec{x}_k, \text{ it follows that}$$

$$T(\vec{y}) = T(a_1\vec{x}_1 + a_2\vec{x}_2 + \ldots + a_k\vec{x}_k) = a_1T(\vec{x}_1) + a_2T(\vec{x}_2) + \ldots + a_kT(\vec{x}_k).$$

These properties are useful in determining the action of a transformation on a given vector. Consider the following example.

Example 4.3.1: 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 9.6.1, 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:

$$\left[\begin{array}{cc|c} 1 & 4 & -7 \\ 3 & 0 & 3 \\ 1 & 5 & -9 \end{array}\right] \to \cdots \to \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{array}\right]$$

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

$$\begin{aligned} T \begin{bmatrix} -7 \\ 3 \\ -9 \end{bmatrix} &= T \left(1 \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} + (-2) \begin{bmatrix} 4 \\ 0 \\ 5 \end{bmatrix} \right) \\ &= 1T \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} - 2T \begin{bmatrix} 4 \\ 0 \\ 5 \end{bmatrix} \\ &= \begin{bmatrix} 4 \\ 4 \\ 0 \\ -2 \end{bmatrix} - 2 \begin{bmatrix} 4 \\ 5 \\ -1 \\ 5 \end{bmatrix} \\ &= \begin{bmatrix} -4 \\ -6 \\ 2 \\ -12 \end{bmatrix} \end{aligned}$$

Therefore, $T \begin{bmatrix} -7 \\ 3 \\ -9 \end{bmatrix} = \begin{bmatrix} -4 \\ -6 \\ 2 \\ -12 \end{bmatrix}$.

Suppose two linear transformations act in the same way on $\vec{x}$ for all vectors. Then we say that these transformations are equal.

Definition 4.3.1: 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 4.3.2: 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 4.3.2: 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 4.3.2: Composition of Transformations

Let T be a linear transformation induced by the matrix

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

and S a linear transformation induced by the matrix

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

Find the matrix of the composite transformation $S \circ T$. Then, find $(S \circ T)(\vec{x})$ for $\vec{x} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$.

Solution

By Theorem 4.3.2, the matrix of $S \circ T$ is given by BA.

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

To find $(S \circ T)(\vec{x})$, multiply $\vec{x}$ by BA as follows

$$\begin{bmatrix} 8 & 4 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 24 \\ 2 \end{bmatrix}$$

To check, first determine $T(\vec{x})$:

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

Then, compute $S(T(\vec{x}))$ as follows:

$$\begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 9 \\ 2 \end{bmatrix} = \begin{bmatrix} 24 \\ 2 \end{bmatrix}$$

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 4.3.3: 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 4.3.3: 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 4.3.3: 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}.

This page titled 4.3: Properties of Linear Transformations. Invertible Linear Transformations. is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **5.3: Properties of Linear Transformations** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.4: Special Linear Transformations and Applications to Computer Graphics

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.

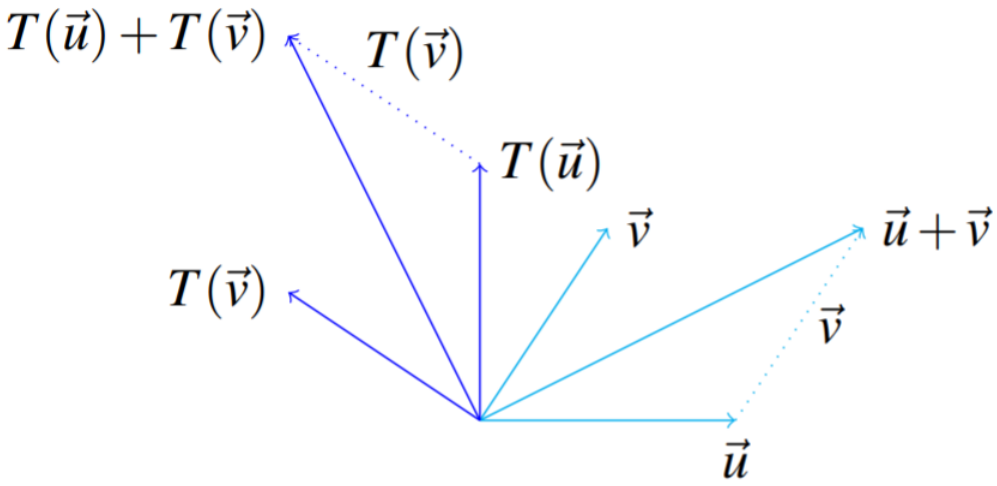


Figure 4.4.1

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 9.6.1.

The following theorem gives the matrix of a linear transformation which rotates all vectors through an angle of θ.

Theorem 4.4.1: 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

$$\begin{bmatrix} \cos[\theta] & -\sin[\theta] \\ \sin[\theta] & \cos[\theta] \end{bmatrix}$$

Proof

Let $\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. These identify the geometric vectors which point along the positive x axis and positive y axis as shown.

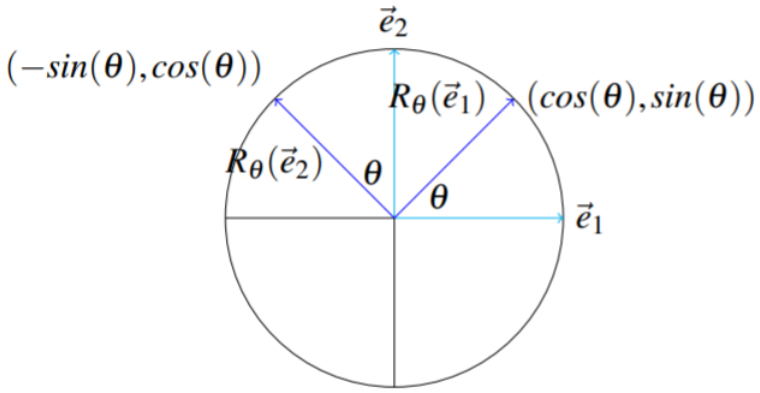


Figure 4.4.2

From Theorem 5.2.2, 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

$$R_\theta(\vec{e}_1) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, R_\theta(\vec{e}_2) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$$

Therefore, from Theorem 5.2.2,

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

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 4.4.1: 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 $\vec{x} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$.

Solution

By Theorem 4.4.1, the matrix of $R_{\frac{\pi}{2}}$ is given by

$$\begin{bmatrix} \cos[\theta] & -\sin[\theta] \\ \sin[\theta] & \cos[\theta] \end{bmatrix} = \begin{bmatrix} \cos[\pi/2] & -\sin[\pi/2] \\ \sin[\pi/2] & \cos[\pi/2] \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

To find $R_{\frac{\pi}{2}}(\vec{x})$, we multiply the matrix of $R_{\frac{\pi}{2}}$ by $\vec{x}$ as follows

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

We now look at an example of a linear transformation involving two angles.

Example 4.4.2: 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 $\vec{u}$

$$R_{\theta+\phi}\left[\vec{u}\right] = A_{\theta+\phi}\vec{u} = A_\theta A_\phi \vec{u} = R_\theta R_\phi\left[\vec{u}\right]$$

Notice the order of the matrices here!

Consequently, you must have

$$\begin{aligned} A_{\theta+\phi} &= \begin{bmatrix} \cos[\theta+\phi] & -\sin[\theta+\phi] \\ \sin[\theta+\phi] & \cos[\theta+\phi] \end{bmatrix} \\ &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} = A_\theta A_\phi \end{aligned}$$

The usual matrix multiplication yields

$$\begin{aligned} A_{\theta+\phi} &= \begin{bmatrix} \cos[\theta+\phi] & -\sin[\theta+\phi] \\ \sin[\theta+\phi] & \cos[\theta+\phi] \end{bmatrix} \\ &= \begin{bmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi & -\cos\theta\sin\phi - \sin\theta\cos\phi \\ \sin\theta\cos\phi + \cos\theta\sin\phi & \cos\theta\cos\phi - \sin\theta\sin\phi \end{bmatrix} \\ &= A_\theta A_\phi \end{aligned}$$

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 4.4.2: 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

$$\frac{1}{1+m^2}\begin{bmatrix} 1-m^2 & 2m \\ 2m & m^2-1 \end{bmatrix}$$

Consider the following example.

Example 4.4.3: 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 $\vec{x} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$.

Solution

By Theorem 4.4.2, the matrix of Q_2 is given by

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

To find $Q_2(\vec{x})$ we multiply $\vec{x}$ by the matrix of Q_2 as follows:

$$\frac{1}{5}\begin{bmatrix} -3 & 8 \\ 8 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -\frac{19}{5} \\ \frac{2}{5} \end{bmatrix}$$

Consider the following example which incorporates a reflection as well as a rotation of vectors.

Example 4.4.4: 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 4.4.1, the matrix of the transformation which involves rotating through an angle of $\pi/6$ is

$$\begin{bmatrix} \cos[\pi/6] & -\sin[\pi/6] \\ \sin[\pi/6] & \cos[\pi/6] \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}\sqrt{3} \end{bmatrix}$$

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 4.4.2, the matrix for the transformation which reflects all vectors through the x axis is

$$\frac{1}{1+m^2}\begin{bmatrix} 1-m^2 & 2m \\ 2m & m^2-1 \end{bmatrix} = \frac{1}{1+(0)^2}\begin{bmatrix} 1-(0)^2 & 2(0) \\ 2(0) & (0)^2-1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Therefore, the matrix of the linear transformation which first rotates through $\pi/6$ and then reflects through the x axis is given by

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

This page titled 4.4: Special Linear Transformations and Applications to Computer Graphics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **5.4: Special Linear Transformations in $\mathbb{R}^2$** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.5: 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 4.5.1: 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 4.5.1: 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 4.5.2: 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 4.5.1: 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.

Note that this proposition says that if $A = \begin{bmatrix} A_1 & \cdots & A_n \end{bmatrix}$ 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 4.5.1: A One to One and Onto Linear Transformation

Suppose

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

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 $\begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2$. Does there exist $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$ such that $T\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$? If so, then since $\begin{bmatrix} a \\ b \end{bmatrix}$ 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

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

This is the same thing as asking for a solution to the following system of equations.

$$\begin{aligned} x + y &= a \\ x + 2y &= b \end{aligned}$$

Set up the augmented matrix and row reduce.

$$\left[\begin{array}{cc|c} 1 & 1 & a \\ 1 & 2 & b \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 2a-b \\ 0 & 1 & b-a \end{array}\right] \tag{4.5.1}$$

You can see from this point that the system has a solution. Therefore, we have shown that for any a, b, there is a $\begin{bmatrix} x \\ y \end{bmatrix}$ such that $T\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$. Thus T is onto.

Now we want to know if T is one to one. By Proposition 4.5.1 it is enough to show that $A\vec{x} = 0$ implies $\vec{x} = 0$. Consider the system $A\vec{x} = 0$ given by:

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

This is the same as the system given by

$$\begin{aligned} x + y &= 0 \\ x + 2y &= 0 \end{aligned}$$

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

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

then

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

In other words, $A\vec{x} = 0$ implies that $\vec{x} = 0$. By Proposition 4.5.1, 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 (4.5.1), 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 $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2a - b \\ b - a \end{bmatrix}$ such that $T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$. Hence by Definition 4.5.1, T is one to one.

Example 4.5.2: 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

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 4.5.1 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.

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 4.5.4: 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.

The next example shows the same concept with regards to one-to-one transformations.

Example 4.5.5: 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.

This page titled 4.5: One-to-One and Onto Transformations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **5.5: One-to-One and Onto Transformations** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.6: Isomorphisms

Outcomes

A. Determine if a linear transformation is an isomorphism.
B. Determine if two subspaces of $\mathbb{R}^n$ are isomorphic.

Recall the definition of a linear transformation. Let V and W be two subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. A mapping $T : V \to W$ is called a **linear transformation** or **linear map** if it preserves the algebraic operations of addition and scalar multiplication. Specifically, if a, b are scalars and $\vec{x}, \vec{y}$ are vectors,

$$T\left(a\vec{x} + b\vec{y}\right) = aT(\vec{x}) + bT(\vec{y})$$

Consider the following important definition.

Definition 4.6.1: Isomorphism

A linear map T is called an **isomorphism** if the following two conditions are satisfied.

- T is one to one. That is, if $T(\vec{x}) = T(\vec{y}),$ then $\vec{x} = \vec{y}$.
- T is onto. That is, if $\vec{w} \in W$, there exists $\vec{v} \in V$ such that $T(\vec{v}) = \vec{w}$.

Two such subspaces which have an isomorphism as described above are said to be **isomorphic.**

Consider the following example of an isomorphism.

Example 4.6.1: Isomorphism

Let $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be defined by

$$T\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x+y \\ x-y \end{bmatrix}$$

Show that T is an isomorphism.

Solution

To prove that T is an isomorphism we must show

1. T is a linear transformation;
2. T is one to one;
3. T is onto.

We proceed as follows.

1. T is a linear transformation:

   Let k, p be scalars.

$$\begin{aligned} T\left(k\begin{bmatrix}x_1\\y_1\end{bmatrix}+p\begin{bmatrix}x_2\\y_2\end{bmatrix}\right) &= T\left(\begin{bmatrix}kx_1\\ky_1\end{bmatrix}+\begin{bmatrix}px_2\\py_2\end{bmatrix}\right)\\ &= T\left(\begin{bmatrix}kx_1+px_2\\ky_1+py_2\end{bmatrix}\right)\\ &= \begin{bmatrix}(kx_1+px_2)+(ky_1+py_2)\\(kx_1+px_2)-(ky_1+py_2)\end{bmatrix}\\ &= \begin{bmatrix}(kx_1+ky_1)+(px_2+py_2)\\(kx_1-ky_1)+(px_2-py_2)\end{bmatrix}\\ &= \begin{bmatrix}kx_1+ky_1\\kx_1-ky_1\end{bmatrix}+\begin{bmatrix}px_2+py_2\\px_2-py_2\end{bmatrix}\\ &= k\begin{bmatrix}x_1+y_1\\x_1-y_1\end{bmatrix}+p\begin{bmatrix}x_2+y_2\\x_2-y_2\end{bmatrix}\\ &= kT\left(\begin{bmatrix}x_1\\y_1\end{bmatrix}\right)+pT\left(\begin{bmatrix}x_2\\y_2\end{bmatrix}\right)\end{aligned}$$

Therefore T is linear.

2. T is one to one:

We need to show that if $T(\vec{x})=\vec{0}$ for a vector $\vec{x}\in\mathbb{R}^2$, then it follows that $\vec{x}=\vec{0}$. Let $\vec{x}=\begin{bmatrix}x\\y\end{bmatrix}$.

$$T\left(\begin{bmatrix}x\\y\end{bmatrix}\right)=\begin{bmatrix}x+y\\x-y\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

This provides a system of equations given by

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

You can verify that the solution to this system if $x=y=0$. Therefore

$$\vec{x}=\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

and T is one to one.

3. T is onto:

Let a,b be scalars. We want to check if there is always a solution to

$$T\left(\begin{bmatrix}x\\y\end{bmatrix}\right)=\begin{bmatrix}x+y\\x-y\end{bmatrix}=\begin{bmatrix}a\\b\end{bmatrix}$$

This can be represented as the system of equations

$$\begin{aligned} x+y&=a\\ x-y&=b\end{aligned}$$

Setting up the augmented matrix and row reducing gives

$$\left[\begin{array}{cc|c}1&1&a\\1&-1&b\end{array}\right]\rightarrow\cdots\rightarrow\left[\begin{array}{cc|c}1&0&\frac{a+b}{2}\\0&1&\frac{a-b}{2}\end{array}\right]$$

This has a solution for all a,b and therefore T is onto.

Therefore T is an isomorphism.

An important property of isomorphisms is that its inverse is also an isomorphism.

Proposition 4.6.1: Inverse of an Isomorphism

Let $T : V \to W$ be an isomorphism and V, W be subspaces of $\mathbb{R}^n$. Then $T^{-1} : W \to V$ is also an isomorphism.

Proof

Let T be an isomorphism. Since T is onto, a typical vector in W is of the form $T(\vec{v})$ where $\vec{v} \in V$. Consider then for a, b scalars,

$$T^{-1}\left(aT(\vec{v}_1) + bT(\vec{v}_2)\right)$$

where $\vec{v}_1, \vec{v}_2 \in V$. Is this equal to

$$aT^{-1}\left(T(\vec{v}_1)\right) + bT^{-1}\left(T(\vec{v}_2)\right) = a\vec{v}_1 + b\vec{v}_2?$$

Since T is one to one, this will be so if

$$T\left(a\vec{v}_1 + b\vec{v}_2\right) = T\left(T^{-1}\left(aT(\vec{v}_1) + bT(\vec{v}_2)\right)\right) = aT(\vec{v}_1) + bT(\vec{v}_2).$$

However, the above statement is just the condition that T is a linear map. Thus T^{-1} is indeed a linear map. If $\vec{v} \in V$ is given, then $\vec{v} = T^{-1}\left(T(\vec{v})\right)$ and so T^{-1} is onto. If $T^{-1}(\vec{v}) = 0$, then

$$\vec{v} = T\left(T^{-1}(\vec{v})\right) = T(\vec{0}) = \vec{0}$$

and so T^{-1} is one to one.

Another important result is that the composition of multiple isomorphisms is also an isomorphism.

Proposition 4.6.2: Composition of Isomorphisms

Let $T : V \to W$ and $S : W \to Z$ be isomorphisms where V, W, Z are subspaces of $\mathbb{R}^n$. Then $S \circ T$ defined by $(S \circ T)(\vec{v}) = S\left(T(\vec{v})\right)$ is also an isomorphism.

Proof

Suppose $T : V \to W$ and $S : W \to Z$ are isomorphisms. Why is $S \circ T$ a linear map? For a, b scalars,

$$\begin{aligned} S \circ T\left(a\vec{v}_1 + b(\vec{v}_2)\right) &= S\left(T\left(a\vec{v}_1 + b\vec{v}_2\right)\right) = S\left(aT\vec{v}_1 + bT\vec{v}_2\right) \\ &= aS\left(T\vec{v}_1\right) + bS\left(T\vec{v}_2\right) = a\left(S \circ T\right)(\vec{v}_1) + b\left(S \circ T\right)(\vec{v}_2) \end{aligned}$$

Hence $S \circ T$ is a linear map. If $(S \circ T)(\vec{v}) = 0$, then $S\left(T(\vec{v})\right) = 0$ and it follows that $T(\vec{v}) = \vec{0}$ and hence by this lemma again, $\vec{v} = \vec{0}$. Thus $S \circ T$ is one to one. It remains to verify that it is onto. Let $\vec{z} \in Z$. Then since S is onto, there exists $\vec{w} \in W$ such that $S(\vec{w}) = \vec{z}$. Also, since T is onto, there exists $\vec{v} \in V$ such that $T(\vec{v}) = \vec{w}$. It follows that $S\left(T(\vec{v})\right) = \vec{z}$ and so $S \circ T$ is also onto.

Consider two subspaces V and W, and suppose there exists an isomorphism mapping one to the other. In this way the two subspaces are related, which we can write as $V \sim W$. Then the previous two propositions together claim that $\sim$ is an equivalence relation. That is: $\sim$ satisfies the following conditions:

- $V \sim V$
- If $V \sim W$, it follows that $W \sim V$
- If $V \sim W$ and $W \sim Z$, then $V \sim Z$

We leave the verification of these conditions as an exercise.

Consider the following example.

Example 4.6.2: Matrix Isomorphism

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be defined by $T(\vec{x}) = A(\vec{x})$ where A is an invertible $n \times n$ matrix. Then T is an isomorphism.

Solution

The reason for this is that, since A is invertible, the only vector it sends to $\vec{0}$ is the zero vector. Hence if $A(\vec{x}) = A(\vec{y})$, then $A(\vec{x} - \vec{y}) = \vec{0}$ and so $\vec{x} = \vec{y}$. It is onto because if

$$\vec{y} \in \mathbb{R}^n, A\left(A^{-1}(\vec{y})\right) = \left(AA^{-1}\right)(\vec{y}) = \vec{y}.$$

In fact, all isomorphisms from $\mathbb{R}^n$ to $\mathbb{R}^n$ can be expressed as $T(\vec{x}) = A(\vec{x})$ where A is an invertible $n \times n$ matrix. One simply considers the matrix whose i^{th} column is $T\vec{e}_i$.

Recall that a basis of a subspace V is a set of linearly independent vectors which span V. The following fundamental lemma describes the relation between bases and isomorphisms.

Lemma 4.6.1: Mapping Bases

Let $T : V \to W$ be a linear transformation where V, W are subspaces of $\mathbb{R}^n$. If T is one to one, then it has the property that if $\{\vec{u}_1, \cdots, \vec{u}_k\}$ is linearly independent, so is $\{T(\vec{u}_1), \cdots, T(\vec{u}_k)\}$.

More generally, T is an isomorphism if and only if whenever $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is a basis for V, it follows that $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ is a basis for W.

Proof

First suppose that T is a linear transformation and is one to one and $\{\vec{u}_1, \cdots, \vec{u}_k\}$ is linearly independent. It is required to show that $\{T(\vec{u}_1), \cdots, T(\vec{u}_k)\}$ is also linearly independent. Suppose then that

$$\sum_{i=1}^{k} c_i T(\vec{u}_i) = \vec{0}$$

Then, since T is linear,

$$T\left(\sum_{i=1}^{n} c_i \vec{u}_i\right) = \vec{0}$$

Since T is one to one, it follows that

$$\sum_{i=1}^{n} c_i \vec{u}_i = 0$$

Now the fact that $\{\vec{u}_1, \cdots, \vec{u}_n\}$ is linearly independent implies that each $c_i = 0$. Hence $\{T(\vec{u}_1), \cdots, T(\vec{u}_n)\}$ is linearly independent.

Now suppose that T is an isomorphism and $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is a basis for V. It was just shown that $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ is linearly independent. It remains to verify that $\text{span}\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\} = W$. If $\vec{w} \in W$, then since T is onto there exists $\vec{v} \in V$ such that $T(\vec{v}) = \vec{w}$. Since $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is a basis, it follows that there exists scalars $\{c_i\}_{i=1}^n$ such that

$$\sum_{i=1}^{n} c_i \vec{v}_i = \vec{v}.$$

Hence,

$$\vec{w} = T(\vec{v}) = T\left(\sum_{i=1}^{n} c_i \vec{v}_i\right) = \sum_{i=1}^{n} c_i T(\vec{v}_i)$$

It follows that $\text{span}\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\} = W$ showing that this set of vectors is a basis for W.

Next suppose that T is a linear transformation which takes a basis to a basis. This means that if $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is a basis for V, it follows $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ is a basis for W. Then if $w \in W$, there exist scalars c_i such that $w = \sum_{i=1}^n c_i T(\vec{v}_i) = T\left(\sum_{i=1}^n c_i \vec{v}_i\right)$ showing that T is onto. If $T\left(\sum_{i=1}^n c_i \vec{v}_i\right) = \vec{0}$ then $\sum_{i=1}^n c_i T(\vec{v}_i) = \vec{0}$ and since the vectors $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ are linearly independent, it follows that each $c_i = 0$. Since $\sum_{i=1}^n c_i \vec{v}_i$ is a typical vector in V, this has shown that if $T(\vec{v}) = \vec{0}$ then $\vec{v} = \vec{0}$ and so T is also one to one. Thus T is an isomorphism.

The following theorem illustrates a very useful idea for defining an isomorphism. Basically, if you know what it does to a basis, then you can construct the isomorphism.

Theorem 4.6.1: Isomorphic Subspaces

Suppose V and W are two subspaces of $\mathbb{R}^n$. Then the two subspaces are isomorphic if and only if they have the same dimension. In the case that the two subspaces have the same dimension, then for a linear map $T : V \to W$, the following are equivalent.

1. T is one to one.
2. T is onto.
3. T is an isomorphism.

Proof

Suppose first that these two subspaces have the same dimension. Let a basis for V be $\{\vec{v}_1, \cdots, \vec{v}_n\}$ and let a basis for W be $\{\vec{w}_1, \cdots, \vec{w}_n\}$. Now define T as follows.

$$T(\vec{v}_i) = \vec{w}_i$$

for $\sum_{i=1}^n c_i \vec{v}_i$ an arbitrary vector of V,

$$T\left(\sum_{i=1}^n c_i \vec{v}_i\right) = \sum_{i=1}^n c_i T\vec{v}_i = \sum_{i=1}^n c_i \vec{w}_i.$$

It is necessary to verify that this is well defined. Suppose then that

$$\sum_{i=1}^n c_i \vec{v}_i = \sum_{i=1}^n \hat{c}_i \vec{v}_i$$

Then

$$\sum_{i=1}^n (c_i - \hat{c}_i)\, \vec{v}_i = \vec{0}$$

and since $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is a basis, $c_i = \hat{c}_i$ for each i. Hence

$$\sum_{i=1}^n c_i \vec{w}_i = \sum_{i=1}^n \hat{c}_i \vec{w}_i$$

and so the mapping is well defined. Also if a, b are scalars,

$$\begin{aligned} T\left(a\sum_{i=1}^n c_i \vec{v}_i + b\sum_{i=1}^n \hat{c}_i \vec{v}_i\right) &= T\left(\sum_{i=1}^n (ac_i + b\hat{c}_i)\, \vec{v}_i\right) = \sum_{i=1}^n (ac_i + b\hat{c}_i)\, \vec{w}_i \\ &= a\sum_{i=1}^n c_i \vec{w}_i + b\sum_{i=1}^n \hat{c}_i \vec{w}_i \\ &= aT\left(\sum_{i=1}^n c_i \vec{v}_i\right) + bT\left(\sum_{i=1}^n \hat{c}_i \vec{v}_i\right) \end{aligned}$$

Thus T is a linear transformation.

Now if

$$T\left(\sum_{i=1}^{n} c_i \vec{v}_i\right) = \sum_{i=1}^{n} c_i \vec{w}_i = \vec{0},$$

then since the $\{\vec{w}_1, \cdots, \vec{w}_n\}$ are independent, each $c_i = 0$ and so $\sum_{i=1}^{n} c_i \vec{v}_i = \vec{0}$ also. Hence T is one to one. If $\sum_{i=1}^{n} c_i \vec{w}_i$ is a vector in W, then it equals

$$\sum_{i=1}^{n} c_i T(\vec{v}_i) = T\left(\sum_{i=1}^{n} c_i \vec{v}_i\right)$$

showing that T is also onto. Hence T is an isomorphism and so V and W are isomorphic.

Next suppose $T : V \mapsto W$ is an isomorphism, so these two subspaces are isomorphic. Then for $\{\vec{v}_1, \cdots, \vec{v}_n\}$ a basis for V, it follows that a basis for W is $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ showing that the two subspaces have the same dimension.

Now suppose the two subspaces have the same dimension. Consider the three claimed equivalences.

First consider the claim that $1.) \Rightarrow 2.)$. If T is one to one and if $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is a basis for V, then $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ is linearly independent. If it is not a basis, then it must fail to span W. But then there would exist $\vec{w} \notin \text{span}\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ and it follows that $\{T(\vec{v}_1), \cdots, T(\vec{v}_n), \vec{w}\}$ would be linearly independent which is impossible because there exists a basis for W of n vectors.

Hence $\text{span}\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\} = W$ and so $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ is a basis. If $\vec{w} \in W$, there exist scalars c_i such that

$$\vec{w} = \sum_{i=1}^{n} c_i T(\vec{v}_i) = T\left(\sum_{i=1}^{n} c_i \vec{v}_i\right)$$

showing that T is onto. This shows that $1.) \Rightarrow 2.)$.

Next consider the claim that $2.) \Rightarrow 3.)$. Since $2.)$ holds, it follows that T is onto. It remains to verify that T is one to one. Since T is onto, there exists a basis of the form $\{T(\vec{v}_i), \cdots, T(\vec{v}_n)\}$. Then it follows that $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is linearly independent. Suppose

$$\sum_{i=1}^{n} c_i \vec{v}_i = \vec{0}$$

Then

$$\sum_{i=1}^{n} c_i T(\vec{v}_i) = \vec{0}$$

Hence each $c_i = 0$ and so, $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is a basis for V. Now it follows that a typical vector in V is of the form $\sum_{i=1}^{n} c_i \vec{v}_i$. If $T\left(\sum_{i=1}^{n} c_i \vec{v}_i\right) = \vec{0}$, it follows that

$$\sum_{i=1}^{n} c_i T(\vec{v}_i) = \vec{0}$$

and so, since $\{T(\vec{v}_i), \cdots, T(\vec{v}_n)\}$ is independent, it follows each $c_i = 0$ and hence $\sum_{i=1}^{n} c_i \vec{v}_i = \vec{0}$. Thus T is one to one as well as onto and so it is an isomorphism.

If T is an isomorphism, it is both one to one and onto by definition so $3.)$ implies both $1.)$ and $2.)$.

Note the interesting way of defining a linear transformation in the first part of the argument by describing what it does to a basis and then "extending it linearly" to the entire subspace.

Example 4.6.4: Isomorphic Subspaces

Let $V = \mathbb{R}^3$ and let W denote

$$\text{span}\left\{\begin{bmatrix}1\\2\\1\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\2\\0\end{bmatrix}\right\}$$

Show that V and W are isomorphic.

Solution

First observe that these subspaces are both of dimension 3 and so they are isomorphic by Theorem 4.6.1. The three vectors which span W are easily seen to be linearly independent by making them the columns of a matrix and row reducing to the reduced row-echelon form.

You can exhibit an isomorphism of these two spaces as follows.

$$T(\vec{e}_1) = \begin{bmatrix}1\\2\\1\\1\end{bmatrix}, T(\vec{e}_2) = \begin{bmatrix}0\\1\\0\\1\end{bmatrix}, T(\vec{e}_3) = \begin{bmatrix}1\\1\\2\\0\end{bmatrix}$$

and extend linearly. Recall that the matrix of this linear transformation is just the matrix having these vectors as columns. Thus the matrix of this isomorphism is

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

You should check that multiplication on the left by this matrix does reproduce the claimed effect resulting from an application by T.

Consider the following example.

Example 4.6.5: Finding the Matrix of an Isomorphism

Let $V = \mathbb{R}^3$ and let W denote

$$\text{span}\left\{\begin{bmatrix}1\\2\\1\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\2\\0\end{bmatrix}\right\}$$

Let $T : V \mapsto W$ be defined as follows.

$$T\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}1\\2\\1\\1\end{bmatrix}, T\begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}0\\1\\0\\1\end{bmatrix}, T\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}1\\1\\2\\0\end{bmatrix}$$

Find the matrix of this isomorphism T.

Solution

First note that the vectors

$$\left[\begin{array}{c}1\\1\\0\end{array}\right], \left[\begin{array}{c}0\\1\\1\end{array}\right], \left[\begin{array}{c}1\\1\\1\end{array}\right]$$

are indeed a basis for $\mathbb{R}^3$ as can be seen by making them the columns of a matrix and using the reduced row-echelon form.

Now recall the matrix of T is a 4×3 matrix A which gives the same effect as T. Thus, from the way we multiply matrices,

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

Hence,

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

Note how the span of the columns of this new matrix must be the same as the span of the vectors defining W.

This idea of defining a linear transformation by what it does on a basis works for linear maps which are not necessarily isomorphisms.

Example 4.6.6: Finding the Matrix of an Isomorphism

Let $V=\mathbb{R}^3$ and let W denote

$$\text{span}\left\{\left[\begin{array}{c}1\\0\\1\\1\end{array}\right], \left[\begin{array}{c}0\\1\\0\\1\end{array}\right], \left[\begin{array}{c}1\\1\\1\\2\end{array}\right]\right\}$$

Let $T: V\mapsto W$ be defined as follows.

$$T\left[\begin{array}{c}1\\1\\0\end{array}\right] = \left[\begin{array}{c}1\\0\\1\\1\end{array}\right], T\left[\begin{array}{c}0\\1\\1\end{array}\right] = \left[\begin{array}{c}0\\1\\0\\1\end{array}\right], T\left[\begin{array}{c}1\\1\\1\end{array}\right] = \left[\begin{array}{c}1\\1\\1\\2\end{array}\right]$$

Find the matrix of this linear transformation.

Solution

Note that in this case, the three vectors which span W are not linearly independent. Nevertheless the above procedure will still work. The reasoning is the same as before. If A is this matrix, then

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

and so

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

> The columns of this last matrix are obviously not linearly independent.

This page titled 4.6: Isomorphisms is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **5.6: Isomorphisms** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.7: The Kernel and Image of A Linear Map

Outcomes

A. Describe the kernel and image of a linear transformation, and find a basis for each.

In this section we will consider the case where the linear transformation is not necessarily an isomorphism. First consider the following important definition.

Definition 4.7.1: Kernel and Image

Let V and W be subspaces of $\mathbb{R}^n$ and let $T : V \mapsto W$ be a linear transformation. Then the image of T denoted as $\text{im}\,(T)$ is defined to be the set

$$\text{im}\,(T) = \{T(\vec{v}) : \vec{v} \in V\}$$

In words, it consists of all vectors in W which equal $T(\vec{v})$ for some $\vec{v} \in V$.

The kernel of T, written $\ker(T)$, consists of all $\vec{v} \in V$ such that $T(\vec{v}) = \vec{0}$. That is,

$$\ker(T) = \left\{\vec{v} \in V : T(\vec{v}) = \vec{0}\right\}$$

It follows that $\text{im}\,(T)$ and $\ker(T)$ are subspaces of W and V respectively.

Proposition 4.7.1: Kernel and Image as Subspaces

Let V, W be subspaces of $\mathbb{R}^n$ and let $T : V \to W$ be a linear transformation. Then $\ker(T)$ is a subspace of V and $\text{im}\,(T)$ is a subspace of W.

Proof

First consider $\ker(T)$. It is necessary to show that if $\vec{v}_1, \vec{v}_2$ are vectors in $\ker(T)$ and if a, b are scalars, then $a\vec{v}_1 + b\vec{v}_2$ is also in $\ker(T)$. But

$$T\left(a\vec{v}_1 + b\vec{v}_2\right) = aT(\vec{v}_1) + bT(\vec{v}_2) = a\vec{0} + b\vec{0} = \vec{0}$$

Thus $\ker(T)$ is a subspace of V.

Next suppose $T(\vec{v}_1), T(\vec{v}_2)$ are two vectors in $\text{im}\,(T)$. Then if a, b are scalars,

$$aT(\vec{v}_2) + bT(\vec{v}_2) = T\left(a\vec{v}_1 + b\vec{v}_2\right)$$

and this last vector is in $\text{im}\,(T)$ by definition.

We will now examine how to find the kernel and image of a linear transformation and describe the basis of each.

Example 4.7.1: Kernel and Image of a Linear Transformation

Let $T : \mathbb{R}^4 \mapsto \mathbb{R}^2$ be defined by

$$T\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} a - b \\ c + d \end{bmatrix}$$

Then T is a linear transformation. Find a basis for $\ker(T)$ and $\text{im}(T)$.

Solution

You can verify that T is a linear transformation.

First we will find a basis for $\ker(T)$. To do so, we want to find a way to describe all vectors $\vec{x} \in \mathbb{R}^4$ such that $T(\vec{x}) = \vec{0}$. Let $\vec{x} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$ be such a vector. Then

$$T \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} a-b \\ c+d \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The values of a, b, c, d that make this true are given by solutions to the system

$$\begin{aligned} a-b &= 0 \\ c+d &= 0 \end{aligned}$$

The solution to this system is $a = s, b = s, c = t, d = -t$ where s, t are scalars. We can describe $\ker(T)$ as follows.

$$\ker(T) = \left\{ \begin{bmatrix} s \\ s \\ t \\ -t \end{bmatrix} \right\} = \text{span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \right\}$$

Notice that this set is linearly independent and therefore forms a basis for $\ker(T)$.

We move on to finding a basis for $\text{im}(T)$. We can write the image of T as

$$\text{im}(T) = \left\{ \begin{bmatrix} a-b \\ c+d \end{bmatrix} \right\}$$

We can write this in the form

$$\text{span} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

This set is clearly not linearly independent. By removing unnecessary vectors from the set we can create a linearly independent set with the same span. This gives a basis for $\text{im}(T)$ as

$$\text{im}(T) = \text{span} \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

Recall that a linear transformation T is called one to one if and only if $T(\vec{x}) = \vec{0}$ implies $\vec{x} = \vec{0}$. Using the concept of kernel, we can state this theorem in another way.

Theorem 4.7.1: One to One and Kernel

Let T be a linear transformation where $\ker(T)$ is the kernel of T. Then T is one to one if and only if $\ker(T)$ consists of **only** the zero vector.

A major result is the relation between the dimension of the kernel and dimension of the image of a linear transformation. In the previous example $\ker(T)$ had dimension 2, and $\text{im}(T)$ also had dimension of 2. Is it a coincidence that the dimension of $\mathbb{M}_{22}$ is $4 = 2 + 2$? Consider the following theorem.

Theorem 4.7.2: Dimension of Kernel and Image

Let $T : V \to W$ be a linear transformation where V, W are subspaces of $\mathbb{R}^n$. Suppose the dimension of V is m. Then

$$m = \dim(\ker(T)) + \dim(\text{im}\,(T))$$

Proof

From Proposition 4.7.1, $\text{im}\,(T)$ is a subspace of W. We know that there exists a basis for $\text{im}\,(T)$, $\{T(\vec{v}_1),\cdots,T(\vec{v}_r)\}$. Similarly, there is a basis for $\ker(T)$, $\{\vec{u}_1,\cdots,\vec{u}_s\}$. Then if $\vec{v}\in V$, there exist scalars c_i such that

$$T(\vec{v})=\sum_{i=1}^{r}c_iT(\vec{v}_i)$$

Hence $T\left(\vec{v}-\sum_{i=1}^{r}c_i\vec{v}_i\right)=0$. It follows that $\vec{v}-\sum_{i=1}^{r}c_i\vec{v}_i$ is in $\ker(T)$. Hence there are scalars a_i such that

$$\vec{v}-\sum_{i=1}^{r}c_i\vec{v}_i=\sum_{j=1}^{s}a_j\vec{u}_j$$

Hence $\vec{v}=\sum_{i=1}^{r}c_i\vec{v}_i+\sum_{j=1}^{s}a_j\vec{u}_j$. Since $\vec{v}$ is arbitrary, it follows that

$$V=\text{span}\,\{\vec{u}_1,\cdots,\vec{u}_s,\vec{v}_1,\cdots,\vec{v}_r\}$$

If the vectors $\{\vec{u}_1,\cdots,\vec{u}_s,\vec{v}_1,\cdots,\vec{v}_r\}$ are linearly independent, then it will follow that this set is a basis. Suppose then that

$$\sum_{i=1}^{r}c_i\vec{v}_i+\sum_{j=1}^{s}a_j\vec{u}_j=0$$

Apply T to both sides to obtain

$$\sum_{i=1}^{r}c_iT(\vec{v}_i)+\sum_{j=1}^{s}a_jT(\vec{u})_j=\sum_{i=1}^{r}c_iT(\vec{v}_i)=0$$

Since $\{T(\vec{v}_1),\cdots,T(\vec{v}_r)\}$ is linearly independent, it follows that each $c_i=0$. Hence $\sum_{j=1}^{s}a_j\vec{u}_j=0$ and so, since the $\{\vec{u}_1,\cdots,\vec{u}_s\}$ are linearly independent, it follows that each $a_j=0$ also. Therefore $\{\vec{u}_1,\cdots,\vec{u}_s,\vec{v}_1,\cdots,\vec{v}_r\}$ is a basis for V and so

$$n=s+r=\dim(\ker(T))+\dim(\text{im}\,(T))$$

The above theorem leads to the next corollary.

Corollary 4.7.1

Let $T:V\to W$ be a linear transformation where V,W are subspaces of $\mathbb{R}^n$. Suppose the dimension of V is m. Then

$$\dim(\ker(T))\leq m$$

$$\dim(\text{im}\,(T))\leq m$$

This follows directly from the fact that $n=\dim(\ker(T))+\dim(\text{im}\,(T))$.

Consider the following example.

Example 4.7.2

Let $T:\mathbb{R}^2\to\mathbb{R}^3$ be defined by

$$T(\vec{x})=\begin{bmatrix}1&0\\1&0\\0&1\end{bmatrix}\vec{x}$$

Then $\text{im}\,(T)=V$ is a subspace of $\mathbb{R}^3$ and T is an isomorphism of $\mathbb{R}^2$ and V. Find a 2×3 matrix A such that the restriction of multiplication by A to $V=\text{im}\,(T)$ equals T^{-1}.

Solution

Since the two columns of the above matrix are linearly independent, we conclude that $\dim(\text{im}(T)) = 2$ and therefore $\dim(\ker(T)) = 2 - \dim(\text{im}(T)) = 2 - 2 = 0$ by Theorem 4.7.2. Then by Theorem 4.7.1 it follows that T is one to one.

Thus T is an isomorphism of $\mathbb{R}^2$ and the two dimensional subspace of $\mathbb{R}^3$ which is the span of the columns of the given matrix. Now in particular,

$$T(\vec{e}_1) = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \; T(\vec{e}_2) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Thus

$$T^{-1} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \vec{e}_1, \; T^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \vec{e}_2$$

Extend T^{-1} to all of $\mathbb{R}^3$ by defining

$$T^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \vec{e}_1$$

Notice that the vectors

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

are linearly independent so T^{-1} can be extended linearly to yield a linear transformation defined on $\mathbb{R}^3$. The matrix of T^{-1} denoted as A needs to satisfy

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

and so

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

Note that

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

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

so the restriction to V of matrix multiplication by this matrix yields T^{-1}.

This page titled 4.7: The Kernel and Image of A Linear Map is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **5.7: The Kernel and Image of A Linear Map** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.8: The Matrix of a Linear Transformation II

Outcomes

A. Find the matrix of a linear transformation with respect to general bases.

We begin this section with an important lemma.

Lemma 4.8.1: Mapping of a Basis

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ be an isomorphism. Then T maps any basis of $\mathbb{R}^n$ to another basis for $\mathbb{R}^n$.

Conversely, if $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a linear transformation which maps a basis of $\mathbb{R}^n$ to another basis of $\mathbb{R}^n$, then it is an isomorphism.

Proof

First, suppose $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a linear transformation which is one to one and onto. Let $\{\vec{v}_1, \cdots, \vec{v}_n\}$ be a basis for $\mathbb{R}^n$. We wish to show that $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ is also a basis for $\mathbb{R}^n$.

First consider why it is linearly independent. Suppose $\sum_{k=1}^n a_k T(\vec{v}_k) = \vec{0}$. Then by linearity we have $T\left(\sum_{k=1}^n a_k \vec{v}_k\right) = \vec{0}$ and since T is one to one, it follows that $\sum_{k=1}^n a_k \vec{v}_k = \vec{0}$. This requires that each $a_k = 0$ because $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is independent, and it follows that $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ is linearly independent.

Next take $\vec{w} \in \mathbb{R}^n$. Since T is onto, there exists $\vec{v} \in \mathbb{R}^n$ such that $T(\vec{v}) = \vec{w}$. Since $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is a basis, in particular it is a spanning set and there are scalars b_k such that $T\left(\sum_{k=1}^n b_k \vec{v}_k\right) = T(\vec{v}) = \vec{w}$. Therefore $\vec{w} = \sum_{k=1}^n b_k T(\vec{v}_k)$ which is in the span $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$. Therefore, $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ is a basis as claimed.

Suppose now that $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a linear transformation such that $T(\vec{v}_i) = \vec{w}_i$ where $\{\vec{v}_1, \cdots, \vec{v}_n\}$ and $\{\vec{w}_1, \cdots, \vec{w}_n\}$ are two bases for $\mathbb{R}^n$.

To show that T is one to one, let $T\left(\sum_{k=1}^n c_k \vec{v}_k\right) = \vec{0}$. Then $\sum_{k=1}^n c_k T(\vec{v}_k) = \sum_{k=1}^n c_k \vec{w}_k = \vec{0}$. It follows that each $c_k = 0$ because it is given that $\{\vec{w}_1, \cdots, \vec{w}_n\}$ is linearly independent. Hence $T\left(\sum_{k=1}^n c_k \vec{v}_k\right) = \vec{0}$ implies that $\sum_{k=1}^n c_k \vec{v}_k = \vec{0}$ and so T is one to one.

To show that T is onto, let $\vec{w}$ be an arbitrary vector in $\mathbb{R}^n$. This vector can be written as $\vec{w} = \sum_{k=1}^n d_k \vec{w}_k = \sum_{k=1}^n d_k T(\vec{v}_k) = T\left(\sum_{k=1}^n d_k \vec{v}_k\right)$. Therefore, T is also onto.

Consider now an important definition.

Definition 4.8.1: Coordinate Vector

Let $B = \{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$ be a basis for $\mathbb{R}^n$ and let $\vec{x}$ be an arbitrary vector in $\mathbb{R}^n$. Then $\vec{x}$ is uniquely represented as $\vec{x} = a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n$ for scalars $a_1, \cdots, a_n$.

The **coordinate vector** of $\vec{x}$ with respect to the basis B, written $C_B(\vec{x})$ or $[\vec{x}]_B$, is given by

$$C_B(\vec{x}) = C_B\left(a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n\right) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

Consider the following example.

Example 4.8.1: Coordinate Vector

Let $B = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$ be a basis of $\mathbb{R}^2$ and let $\vec{x} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$ be a vector in $\mathbb{R}^2$. Find $C_B(\vec{x})$.

Solution

First, note the order of the basis is important so label the vectors in the basis B as

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\} = \{\vec{v}_1, \vec{v}_2\}$$

Now we need to find a_1, a_2 such that $\vec{x} = a_1\vec{v}_1 + a_2\vec{v}_2$, that is:

$$\begin{bmatrix} 3 \\ -1 \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Solving this system gives $a_1 = 2, a_2 = -1$. Therefore the coordinate vector of $\vec{x}$ with respect to the basis B is

$$C_B(\vec{x}) = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

Given any basis B, one can easily verify that the coordinate function is actually an isomorphism.

Theorem 4.8.1: C_BTransformation is a Linear

For any basis B of $\mathbb{R}^n$, the coordinate function

$$C_B : \mathbb{R}^n \to \mathbb{R}^n$$

is a linear transformation, and moreover an isomorphism.

We now discuss the main result of this section, that is how to represent a linear transformation with respect to different bases.

Theorem 4.8.2: The Matrix of a Linear

Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear transformation, and let B_1 and B_2 be bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively.

Then the following holds

$$C_{B_2} T = M_{B_2 B_1} C_{B_1} \tag{4.8.1}$$

where $M_{B_2 B_1}$ is a unique $m \times n$ matrix.

If the basis B_1 is given by $B_1 = \{\vec{v}_1, \cdots, \vec{v}_n\}$ in this order, then

$$M_{B_2 B_1} = [C_{B_2}(T(\vec{v}_1)) \; C_{B_2}(T(\vec{v}_2)) \; \cdots C_{B_2}(T(\vec{v}_n))]$$

Proof

The above equation (4.8.1) can be represented by the following diagram.

$$\begin{array}{ccc} & T & \\ \mathbb{R}^n & \to & \mathbb{R}^m \\ C_{B_1} \downarrow & \circ & \downarrow C_{B_2} \\ \mathbb{R}^n & \to & \mathbb{R}^m \\ & M_{B_2 B_1} & \end{array}$$

Since C_{B_1} is an isomorphism, then the matrix we are looking for is the matrix of the linear transformation

$$C_{B_2} T C_{B_1}^{-1} : \mathbb{R}^n \mapsto \mathbb{R}^m.$$

By Theorem 5.2.2, the columns are given by the image of the standard basis $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$. But since $C_{B_1}^{-1}(\vec{e}_i) = \vec{v}_i$, we readily obtain that

$$\begin{aligned} M_{B_2B_1} &= \left[C_{B_2}TC_{B_1}^{-1}(\vec{e}_1) \;\; C_{B_2}TC_{B_1}^{-1}(\vec{2}_2) \;\; \cdots \;\; C_{B_2}TC_{B_1}^{-1}(\vec{e}_n)\right] \\ &= [C_{B_2}(T(\vec{v}_1)) \;\; C_{B_2}(T(\vec{v}_2)) \;\; \cdots \;\; C_{B_2}(T(\vec{v}_n))] \end{aligned}$$

and this completes the proof.

Consider the following example.

Example 4.8.2: Matrix of a Linear

Let $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be a linear transformation defined by $T\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} b \\ a \end{bmatrix}$.

Consider the two bases

$$B_1 = \{\vec{v}_1, \vec{v}_2\} = \left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\}$$

and

$$B_2 = \left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\}$$

Find the matrix M_{B_2,B_1} of T with respect to the bases B_1 and B_2.

Solution

By Theorem 4.8.2, the columns of $M_{B_2B_1}$ are the coordinate vectors of $T(\vec{v}_1), T(\vec{v}_2)$ with respect to B_2.

Since

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

a standard calculation yields

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \left(\frac{1}{2}\right)\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \left(-\frac{1}{2}\right)\begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

the first column of $M_{B_2B_1}$ is $\begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$.

The second column is found in a similar way. We have

$$T\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

and with respect to B_2 calculate:

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

Hence the second column of $M_{B_2B_1}$ is given by $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. We thus obtain

$$M_{B_2B_1} = \begin{bmatrix} \frac{1}{2} & 0 \\ -\frac{1}{2} & 1 \end{bmatrix}$$

We can verify that this is the correct matrix $M_{B_2B_1}$ on the specific example

$$\vec{v} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

First applying T gives

$$T(\vec{v}) = T\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

and one can compute that

$$C_{B_2}\left(\begin{bmatrix} -1 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

On the other hand, one compute $C_{B_1}(\vec{v})$ as

$$C_{B_1}\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ -1 \end{bmatrix},$$

and finally applying $M_{B_1B_2}$ gives

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

as above.

We see that the same vector results from either method, as suggested by Theorem 4.8.2.

If the bases B_1 and B_2 are equal, say B, then we write M_B instead of M_{BB}. The following example illustrates how to compute such a matrix. Note that this is what we did earlier when we considered only $B_1 = B_2$ to be the standard basis.

Example 4.8.3: Matrix of a Linear Transformation with respect to an Arbitrary

Consider the basis B of $\mathbb{R}^3$ given by

$$B = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

And let $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$ be the linear transformation defined on B as:

$$T\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, T\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, T\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

1. Find the matrix M_B of T relative to the basis B.
2. Then find the usual matrix of T with respect to the standard basis of $\mathbb{R}^3$.

Solution

Equation (4.8.1) gives $C_BT = M_BC_B$, and thus $M_B = C_BTC_B^{-1}$.

Now $C_B(\vec{v}_i) = \vec{e}_i$, so the matrix of C_B^{-1} (with respect to the standard basis) is given by

$$\begin{bmatrix} C_B^{-1}(\vec{e}_1) & C_B^{-1}(\vec{e}_2) & C_B^{-1}(\vec{e}_2) \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Moreover the matrix of TC_B^{-1} is given by

$$\begin{bmatrix} TC_B^{-1}(\vec{e}_1) & TC_B^{-1}(\vec{e}_2) & TC_B^{-1}(\vec{e}_2) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 2 & 1 \\ 1 & -1 & 1 \end{bmatrix}$$

Thus

$$\begin{aligned} M_B &= C_B T C_B^{-1} = [C_B^{-1}]^{-1}[TC_B^{-1}] \\ &= \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 0 \\ -1 & 2 & 1 \\ 1 & -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 & -5 & 1 \\ -1 & 4 & 0 \\ 0 & -2 & 1 \end{bmatrix} \end{aligned}$$

Consider how this works. Let $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ be an arbitrary vector in $\mathbb{R}^3$.

Apply C_B^{-1} to $\vec{b}$ to get

$$b_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + b_2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + b_3 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

Apply T to this linear combination to obtain

$$b_1 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} + b_2 \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + b_3 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} b_1 + b_2 \\ -b_1 + 2b_2 + b_3 \\ b_1 - b_2 + b_3 \end{bmatrix}$$

Now take the matrix M_B of the transformation (as found above) and multiply it by $\vec{b}$.

$$\begin{bmatrix} 2 & -5 & 1 \\ -1 & 4 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 2b_1 - 5b_2 + b_3 \\ -b_1 + 4b_2 \\ -2b_2 + b_3 \end{bmatrix}$$

Is this the coordinate vector of the above relative to the given basis? We check as follows.

$$\begin{aligned} &(2b_1 - 5b_2 + b_3) \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + (-b_1 + 4b_2) \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + (-2b_2 + b_3) \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} b_1 + b_2 \\ -b_1 + 2b_2 + b_3 \\ b_1 - b_2 + b_3 \end{bmatrix} \end{aligned}$$

You see it is the same thing.

Now lets find the matrix of T with respect to the standard basis. Let A be this matrix. That is, multiplication by A is the same as doing T. Thus

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

Hence

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

Of course this is a very different matrix than the matrix of the linear transformation with respect to the non standard basis.

This page titled 4.8: The Matrix of a Linear Transformation II is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **5.8: The Matrix of a Linear Transformation II** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.9: 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 4.9.1: 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 4.9.2: Null Space or Kernel of a Linear Transformation

Let T be a linear transformation. Define

$$\ker(T)=\left\{\vec{x}:T\left(\vec{x}\right)=\vec{0}\right\}$$

The kernel, $\ker(T)$ 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 4.9.1: 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 4.9.2 states that $\ker(T)$ is the set of solutions to the equation,

$$T\left(\vec{x}\right)=\vec{0}$$

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 4.9.1: 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 4.9.2: The Null Space of a Matrix

Let

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

Find $\text{null}(A)$. 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

$$\begin{aligned} x+2y+3z &= 0 \\ 2x+y+z+2w &= 0 \\ 4x+5y+7z+2w &= 0 \end{aligned}$$

To solve, set up the augmented matrix and row reduce to find the reduced row-echelon form.

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & 0 & 0 \\ 2 & 1 & 1 & 2 & 0 \\ 4 & 5 & 7 & 2 & 0 \end{array}\right] \to \cdots \to \left[\begin{array}{cccc|c} 1 & 0 & -\frac{1}{3} & \frac{4}{3} & 0 \\ 0 & 1 & \frac{5}{3} & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

This yields $x = \frac{1}{3}z - \frac{4}{3}w$ and $y = \frac{2}{3}w - \frac{5}{3}z$. Since null (A) consists of the solutions to this system, it consists vectors of the form,

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

Consider the following example.

Example 4.9.3: 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 4.9.2. Therefore, from Theorem 4.9.1, 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

$$\vec{x}_p = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}$$

Using this particular solution along with the solutions found in Example 4.9.2, 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.

This page titled 4.9: The General Solution of a Linear System is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **5.9: The General Solution of a Linear System** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.E: Exercises

Exercise 4.E.1

Show the map T: $\mathbb{R}^n \to \mathbb{R}^m$ 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.

Answer

This result follows from the properties of matrix multiplication.

Exercise 4.E.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.

Answer

$$\begin{aligned} T_{\vec{u}}(a\vec{v}+b\vec{w}) &= a\vec{v}+b\vec{w} - \frac{(a\vec{v}+b\vec{w} \bullet \vec{u})}{||\vec{u}||^2}\vec{u} \\ &= a\vec{v} - a\frac{(\vec{v}\bullet\vec{u})}{||\vec{u}||^2}\vec{u} + b\vec{w} - b\frac{(\vec{w}\bullet\vec{u})}{||\vec{u}||^2}\vec{u} \\ &= aT_{\vec{u}}(\vec{v}) + bT_{\vec{u}}(\vec{w}) \end{aligned}$$

Exercise 4.E.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.

Answer

Linear transformations take $\vec{0}$ to $\vec{0}$ which T does not. Also $T_{\vec{a}}(\vec{u}+\vec{v}) \neq T_{\vec{a}}\vec{u} + T_{\vec{a}}\vec{v}$.

Exercise 4.E.4

Consider the following functions which map $\mathbb{R}^n$ to $\mathbb{R}^n$.

a. T multiplies the jth component of $\vec{x}$ by a nonzero number b.
b. T replaces the ith component of $\vec{x}$ with b times the jth component added to the ith component.
c. T switches the ith and jth components.

Show these functions are linear transformations and describe their matrices A such that $T(\vec{x}) = A\vec{x}$.

Answer

a. The matrix of T is the elementary matrix which multiplies the jth diagonal entry of the identity matrix by b.
b. The matrix of T is the elementary matrix which takes b times the jth row and adds to the ith row.
c. The matrix of T is the elementary matrix which switches the ith and the jth rows where the two components are in the ith and jth positions.

Exercise 4.E.5

You are given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ and you know that

$$T(A_i) = B_i$$

where $[A_1 \cdots a_n]^{-1}$ exists. Show that the matrix of T is of the form

$$[B_1 \cdots B_n][A_1 \cdots A_n]^{-1}$$

Answer

Suppose

$$\begin{bmatrix} \vec{c}_1^T \\ \vdots \\ \vec{c}_n^T \end{bmatrix} = [\,\vec{a}_1 \quad \cdots \quad \vec{a}_n\,]^{-1}$$

Thus $\vec{c}_i^T \vec{a}_j = \delta_{ij}$. Therefore

$$\begin{aligned} [\,\vec{b}_1 \quad \cdots \quad \vec{b}_n\,]\,[\,\vec{a}_1 \quad \cdots \quad \vec{a}_n\,]^{-1}\vec{a}_i &= [\,\vec{b}_1 \quad \cdots \quad \vec{b}_n\,] \begin{bmatrix} \vec{c}_1^T \\ \vdots \\ \vec{c}_n^T \end{bmatrix} \vec{a}_i \\ &= [\,\vec{b}_1 \quad \cdots \quad \vec{b}_n\,]\,\vec{e}_i \\ &= \vec{b}_i \end{aligned}$$

Thus $T\vec{a}_i = [\,\vec{b}_1 \quad \cdots \quad \vec{b}_n\,]\,[\,\vec{a}_1 \quad \cdots \quad \vec{a}_n\,]^{-1}\vec{a}_i = A\vec{a}_i$. If $\vec{x}$ is arbitrary, then since the matrix $[\,\vec{a}_1 \quad \cdots \quad \vec{a}_n\,]$ is invertible, there exists a unique $\vec{y}$ such that $[\,\vec{a}_1 \quad \cdots \quad \vec{a}_n\,]\,\vec{y} = \vec{x}$ Hence

$$T\vec{x} = T\left(\sum_{i=1}^{n} y_i \vec{a}_i\right) = \sum_{i=1}^{n} y_i T\vec{a}_i = \sum_{i=1}^{n} y_i A\vec{a}_i = A\left(\sum_{i=1}^{n} y_i \vec{a}_i\right) = A\vec{x}$$

Exercise $4.E.6$

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 $T(\vec{x}) = A\vec{x}$.

Answer

$$\begin{bmatrix} 5 & 1 & 5 \\ 1 & 1 & 3 \\ 3 & 5 & -2 \end{bmatrix}\begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 4 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 37 & 17 & 11 \\ 17 & 7 & 5 \\ 11 & 14 & 6 \end{bmatrix}$$

Exercise $4.E.7$

Suppose T is a linear transformation such that

https://math.libretexts.org/@go/page/170465

$$T\begin{bmatrix}1\\1\\-8\end{bmatrix}=\begin{bmatrix}1\\3\\1\end{bmatrix}$$
$$T\begin{bmatrix}-1\\0\\6\end{bmatrix}=\begin{bmatrix}2\\4\\1\end{bmatrix}$$
$$T\begin{bmatrix}0\\-1\\3\end{bmatrix}=\begin{bmatrix}6\\1\\-1\end{bmatrix}$$

Find the matrix of T. That is find A such that $T(\vec{x})=A\vec{x}$.

Answer

$$\begin{bmatrix}1&2&6\\3&4&1\\1&1&-1\end{bmatrix}\begin{bmatrix}6&3&1\\5&3&1\\6&2&1\end{bmatrix}=\begin{bmatrix}52&21&9\\44&23&8\\5&4&1\end{bmatrix}$$

Exercise $4.E.8$

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 $T(\vec{x})=A\vec{x}$.

Answer

$$\begin{bmatrix}-3&1&5\\1&3&3\\3&-3&-3\end{bmatrix}\begin{bmatrix}2&2&1\\1&2&1\\4&1&1\end{bmatrix}=\begin{bmatrix}15&1&3\\17&11&7\\-9&-3&-3\end{bmatrix}$$

Exercise $4.E.9$

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 $T(\vec{x}) = A\vec{x}$.

Answer

$$\begin{bmatrix} 3 & 1 & 1 \\ 3 & 2 & 3 \\ 3 & 3 & -1 \end{bmatrix} \begin{bmatrix} 6 & 2 & 1 \\ 5 & 2 & 1 \\ 6 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 29 & 9 & 5 \\ 46 & 13 & 8 \\ 27 & 11 & 5 \end{bmatrix}$$

Exercise $4.E.10$

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 $T(\vec{x}) = A\vec{x}$.

Answer

$$\begin{bmatrix} 5 & 3 & 2 \\ 2 & 3 & 5 \\ 5 & 5 & -2 \end{bmatrix} \begin{bmatrix} 11 & 4 & 1 \\ 10 & 4 & 1 \\ 12 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 109 & 38 & 10 \\ 112 & 35 & 10 \\ 81 & 34 & 8 \end{bmatrix}$$

Exercise $4.E.11$

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 $T(\vec{x}) = A\vec{x}$.

a. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+2y+3z \\ 2y-3x+z \end{bmatrix}$

b. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 7x+2y+z \\ 3x-11y+2z \end{bmatrix}$

c. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3x+2y+z \\ x+2y+6z \end{bmatrix}$

d. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2y-5x+z \\ x+y+z \end{bmatrix}$

Exercise $4.E.12$

Consider the following functions $T : \mathbb{R}^3 \to \mathbb{R}^2$. Explain why each of these functions T is not linear.

https://math.libretexts.org/@go/page/170465

a. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+2y+3z+1 \\ 2y-3x+z \end{bmatrix}$

b. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+2y^2+3z \\ 2y+3x+z \end{bmatrix}$

c. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \sin x+2y+3z \\ 2y+3x+z \end{bmatrix}$

d. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+2y+3z \\ 2y+3x-\ln z \end{bmatrix}$

Exercise 4.E.13

Suppose

$$[A_1 \cdots A_n]^{-1}$$

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 4.E.14

Find the matrix for $T(\vec{w}) = proj_{\vec{v}}(\vec{w})$ where $\vec{v} = [\,1 \quad -2 \quad 3\,]^T$.

Answer

Recall that $proj_{\vec{u}}(\vec{v}) = \frac{\vec{v}\bullet\vec{u}}{||\vec{u}||^2}\vec{u}$ and so the desired matrix has ith column equal to $proj_{\vec{u}}(\vec{e}_i)$. Therefore, the matrix desired is

$$\frac{1}{14}\begin{bmatrix} 1 & -2 & 3 \\ -2 & 4 & -6 \\ 3 & -6 & 9 \end{bmatrix}$$

Exercise 4.E.15

Find the matrix for $T(\vec{w}) = proj_{\vec{v}}(\vec{w})$ where $\vec{v} = [\,1 \quad 5 \quad 3\,]^T$.

Answer

$$\frac{1}{35}\begin{bmatrix} 1 & 5 & 3 \\ 5 & 25 & 15 \\ 3 & 15 & 9 \end{bmatrix}$$

Exercise 4.E.16

Find the matrix for $T(\vec{w}) = proj_{\vec{v}}(\vec{w})$ where $\vec{v} = [\,1 \quad 0 \quad 3\,]^T$.

Answer

$$\frac{1}{10}\begin{bmatrix} 1 & 0 & 3 \\ 0 & 0 & 0 \\ 3 & 0 & 9 \end{bmatrix}$$

Exercise 4.E. 17

Show that if a function $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then it is always the case that $T(\vec{0}) = \vec{0}$.

Exercise 4.E. 18

Let T be a linear transformation induced by the matrix $A = \begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix}$ and S a linear transformation induced by $B = \begin{bmatrix} 0 & -2 \\ 4 & 2 \end{bmatrix}$. Find matrix of $S \circ T$ and find $(S \circ T)(\vec{x})$ for $\vec{x} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

Answer

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, $(S \circ T)(\vec{x}) = (BA)\vec{x}$.

$$\begin{bmatrix} 2 & -4 \\ 10 & 8 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 8 \\ 12 \end{bmatrix}$$

Exercise 4.E. 19

Let T be a linear transformation and suppose $T\left(\begin{bmatrix} 1 \\ -4 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$. Suppose S is a linear transformation induced by the matrix $B = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}$. Find $(S \circ T)(\vec{x})$ for $\vec{x} = \begin{bmatrix} 1 \\ -4 \end{bmatrix}$.

Answer

To find $(S \circ T)(\vec{x})$ we compute $S(T(\vec{x}))$.

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

Exercise 4.E. 20

Let T be a linear transformation induced by the matrix $A = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}$ and S a lienar transformation induced by $B = \begin{bmatrix} -1 & 3 \\ 1 & -2 \end{bmatrix}$. Find matrix of $S \circ T$ and find $(S \circ T)(\vec{x})$ for $\vec{x} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$.

Exercise 4.E. 21

Let T be a linear transformation induced by the matrix $A = \begin{bmatrix} 2 & 1 \\ 5 & 2 \end{bmatrix}$. Find the matrix of T^{-1}.

Answer

The matrix of T^{-1} is A^{-1}.

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

Exercise 4.E.22

Let T be a linear transformation induced by the matrix $A = \begin{bmatrix} 4 & -3 \\ 2 & -2 \end{bmatrix}$. Find the matrix of T^{-1}.

Exercise 4.E.23

Let T be a linear transformation and suppose $T\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$, $T\left(\begin{bmatrix} 0 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} -4 \\ -3 \end{bmatrix}$. Find the matrix of T^{-1}.

Exercise 4.E.24

Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/3$.

Answer

$$\begin{bmatrix} \cos\left(\frac{\pi}{3}\right) & -\sin\left(\frac{\pi}{3}\right) \\ \sin\left(\frac{\pi}{3}\right) & \cos\left(\frac{\pi}{3}\right) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & \frac{1}{2} \end{bmatrix}$$

Exercise 4.E.25

Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/4$.

Answer

$$\begin{bmatrix} \cos\left(\frac{\pi}{4}\right) & -\sin\left(\frac{\pi}{4}\right) \\ \sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{bmatrix}$$

Exercise 4.E.26

Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $-\pi/3$.

Answer

$$\begin{bmatrix} \cos\left(-\frac{\pi}{3}\right) & -\sin\left(-\frac{\pi}{3}\right) \\ \sin\left(-\frac{\pi}{3}\right) & \cos\left(-\frac{\pi}{3}\right) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} & \frac{1}{2} \end{bmatrix}$$

Exercise 4.E.27

Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $2\pi/3$.

Answer

$$\begin{bmatrix} \cos\left(\frac{2\pi}{3}\right) & -\sin\left(\frac{2\pi}{3}\right) \\ \sin\left(\frac{2\pi}{3}\right) & \cos\left(\frac{2\pi}{3}\right) \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{bmatrix}$$

Exercise 4.E.28

Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/12$. Hint: Note that $\pi/12 = \pi/3 - \pi/4$.

Answer

https://math.libretexts.org/@go/page/170465

$$\begin{bmatrix} \cos\left(\frac{\pi}{3}\right) & -\sin\left(\frac{\pi}{3}\right) \\ \sin\left(\frac{\pi}{3}\right) & \cos\left(\frac{\pi}{3}\right) \end{bmatrix} \begin{bmatrix} \cos\left(-\frac{\pi}{4}\right) & -\sin\left(-\frac{\pi}{4}\right) \\ \sin\left(-\frac{\pi}{4}\right) & \cos\left(-\frac{\pi}{4}\right) \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{4}\sqrt{2}\sqrt{3}+\frac{1}{4}\sqrt{2} & \frac{1}{4}\sqrt{2}-\frac{1}{4}\sqrt{2}\sqrt{3} \\ \frac{1}{4}\sqrt{2}\sqrt{3}-\frac{1}{4}\sqrt{2} & \frac{1}{4}\sqrt{2}\sqrt{3}+\frac{1}{4}\sqrt{2} \end{bmatrix}$$

? Exercise 4.E.29

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.

Answer

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos\left(\frac{2\pi}{3}\right) & -\sin\left(\frac{2\pi}{3}\right) \\ \sin\left(\frac{2\pi}{3}\right) & \cos\left(\frac{2\pi}{3}\right) \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} & \frac{1}{2} \end{bmatrix}$$

? Exercise 4.E.30

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.

Answer

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos\left(\frac{\pi}{3}\right) & -\sin\left(\frac{\pi}{3}\right) \\ \sin\left(\frac{\pi}{3}\right) & \cos\left(\frac{\pi}{3}\right) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{bmatrix}$$

? Exercise 4.E.31

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.

Answer

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos\left(\frac{\pi}{4}\right) & -\sin\left(\frac{\pi}{4}\right) \\ \sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \end{bmatrix}$$

? Exercise 4.E.32

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.

Answer

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\left(\frac{\pi}{6}\right) & -\sin\left(\frac{\pi}{6}\right) \\ \sin\left(\frac{\pi}{6}\right) & \cos\left(\frac{\pi}{6}\right) \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}\sqrt{3} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}\sqrt{3} \end{bmatrix}$$

? Exercise 4.E.33

Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ across the x axis and then rotates every vector through an angle of $\pi/4$.

Answer

$$\begin{bmatrix} \cos\left(\frac{\pi}{4}\right) & -\sin\left(\frac{\pi}{4}\right) \\ \sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \end{bmatrix}$$

Exercise $4.E.34$

Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ across the y axis and then rotates every vector through an angle of $\pi/4$.

Answer

$$\begin{bmatrix} \cos\left(\frac{\pi}{4}\right) & -\sin\left(\frac{\pi}{4}\right) \\ \sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{bmatrix}$$

Exercise $4.E.35$

Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ across the x axis and then rotates every vector through an angle of $\pi/6$.

Answer

$$\begin{bmatrix} \cos\left(\frac{\pi}{6}\right) & -\sin\left(\frac{\pi}{6}\right) \\ \sin\left(\frac{\pi}{6}\right) & \cos\left(\frac{\pi}{6}\right) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\sqrt{3} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2}\sqrt{3} \end{bmatrix}$$

Exercise $4.E.36$

Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ across the y axis and then rotates every vector through an angle of $\pi/6$.

Answer

$$\begin{bmatrix} \cos\left(\frac{\pi}{6}\right) & -\sin\left(\frac{\pi}{6}\right) \\ \sin\left(\frac{\pi}{6}\right) & \cos\left(\frac{\pi}{6}\right) \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2}\sqrt{3} \end{bmatrix}$$

Exercise $4.E.37$

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 $5\pi/12 = 2\pi/3 - \pi/4$

Answer

$$\begin{bmatrix} \cos\left(\frac{2\pi}{3}\right) & -\sin\left(\frac{2\pi}{3}\right) \\ \sin\left(\frac{2\pi}{3}\right) & \cos\left(\frac{2\pi}{3}\right) \end{bmatrix} \begin{bmatrix} \cos\left(-\frac{\pi}{4}\right) & -\sin\left(-\frac{\pi}{4}\right) \\ \sin\left(-\frac{\pi}{4}\right) & \cos\left(-\frac{\pi}{4}\right) \end{bmatrix} =$$

$$\begin{bmatrix} \frac{1}{4}\sqrt{2}\sqrt{3} - \frac{1}{4}\sqrt{2} & -\frac{1}{4}\sqrt{2}\sqrt{3} - \frac{1}{4}\sqrt{2} \\ \frac{1}{4}\sqrt{2}\sqrt{3} + \frac{1}{4}\sqrt{2} & \frac{1}{4}\sqrt{2}\sqrt{3} - \frac{1}{4}\sqrt{2} \end{bmatrix}$$

Note that it doesn't matter about the order in this case.

https://math.libretexts.org/@go/page/170465

Exercise 4.E.38

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.

Answer

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \cos\left(\frac{\pi}{6}\right) & -\sin\left(\frac{\pi}{6}\right) & 0 \\ \sin\left(\frac{\pi}{6}\right) & \cos\left(\frac{\pi}{6}\right) & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\sqrt{3} & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2}\sqrt{3} & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Exercise 4.E.39

Let $\vec{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ be a unit vector in $\mathbb{R}^2$. Find the matrix which reflects all vectors across this vector, as shown in the following picture.

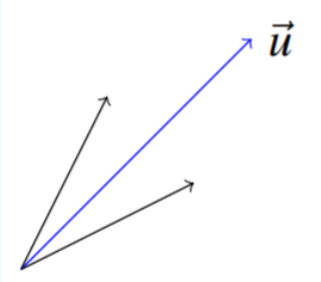


Figure 4.E.1

Hint: Notice that $\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ for some θ. First rotate through $-\theta$. Next reflect through the x axis. Finally rotate through θ.

Answer

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos(-\theta) & -\sin(\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix}$$
$$= \begin{bmatrix} \cos^2\theta - \sin^2\theta & 2\cos\theta\sin\theta \\ 2\cos\theta\sin\theta & \sin^2\theta - \cos^2\theta \end{bmatrix}$$

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

$$\begin{bmatrix} a^2-b^2 & 2ab \\ 2ab & b^2-a^2 \end{bmatrix}$$

https://math.libretexts.org/@go/page/170465

Exercise 4.E.40

Let T be a linear transformation given by

$$T\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Is T one to one? Is T onto?

Exercise 4.E.41

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 4.E.42

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 4.E.43

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 4.E.44

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.

Answer

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

Exercise 4.E.45

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.

Answer

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.

Exercise $4.E.46$

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.

Answer

The columns are independent. Therefore, A is one to one.

Exercise $4.E.47$

Explain why an $n \times n$ matrix A is both one to one and onto if and only if its rank is n.

Answer

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.

Exercise $4.E.48$

Let V and W be subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively and let $T : V \to W$ be a linear transformation. Suppose that $\{T\vec{v}_1, \cdots, T\vec{v}_r\}$ is linearly independent. Show that it must be the case that $\{T\vec{v}_1, \cdots, T\vec{v}_r\}$ is also linearly independent.

Answer

If $\sum_i^r a_i \vec{v}_r = 0$, then using linearity properties of T we get

$$0 = T(0) = T\left(\sum_i^r a_i \vec{v}_r\right) = \sum_i^r a_i T(\vec{v}_r).$$

Since we assume that $\{T\vec{v}_a, \cdots, T\vec{v}_r\}$ is linearly independent, we must have all $a_i = 0$, and therefore we conclude that $\{\vec{v}_1, \cdots, \vec{v}_r\}$ is also linearly independent.

Exercise $4.E.49$

Let

$$V = span\left\{\begin{bmatrix}1\\1\\2\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\0\\1\end{bmatrix}\right\}$$

Let $T\vec{x} = A\vec{x}$ where A is the matrix

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

Give a basis for $im(T)$.

Exercise 4.E.50

Let

$$V = span\left\{\begin{bmatrix}1\\0\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\4\\4\\1\end{bmatrix}\right\}$$

Let $T\vec{x} = A\vec{x}$ where A is the matrix

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

Find a basis for $im(T)$. In this case, the original vectors do not form an independent set.

Answer

Since the third vector is a linear combinations of the first two, then the image of the third vector will also be a linear combinations of the image of the first two. However the image of the first two vectors are linearly independent (check!), and hence form a basis of the image.

Thus a basis for $im(T)$ is:

$$V = span\left\{\begin{bmatrix}2\\0\\1\\3\end{bmatrix}, \begin{bmatrix}4\\2\\4\\5\end{bmatrix}\right\}$$

Exercise 4.E.51

If $\{\vec{v}_1, \cdots, \vec{v}_r\}$ is linearly independent and T is a one to one linear transformation, show that $\{T\vec{v}_1, \cdots, T\vec{v}_r\}$ is also linearly independent. Give an example which shows that if T is only linear, it can happen that, although $\{\vec{v}_1, \cdots, \vec{v}_r\}$ is linearly independent, $\{T\vec{v}_1, \cdots, T\vec{v}_r\}$ is not. In fact, show that it can happen that each of the $T\vec{v}_j$ equals 0.

Exercise 4.E.52

Let V and W be subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively and let $T : V \to W$ be a linear transformation. Show that if T is onto W and if $\{\vec{v}_1, \cdots, \vec{v}_r\}$ is a basis for V, then $span\{T\vec{v}_1, \cdots, T\vec{v}_r\} = W$.

Exercise 4.E.53

Define T: $\mathbb{R}^4 \to \mathbb{R}^3$ as follows.

$$T\vec{x} = \begin{bmatrix}3 & 2 & 1 & 8\\2 & 2 & -2 & 6\\1 & 1 & -1 & 3\end{bmatrix}\vec{x}$$

Find a basis for $im(T)$. Also find a basis for $\ker(T)$.

Exercise 4.E.54

Define $T: \mathbb{R}^3 \to \mathbb{R}^3$ as follows.

$$T\vec{x} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \vec{x}$$

where on the right, it is just matrix multiplication of the vector $\vec{x}$ which is meant. Explain why T is an isomorphism of $\mathbb{R}^3$ to $\mathbb{R}^3$.

Exercise 4.E.55

Suppose $T: \mathbb{R}^3 \to \mathbb{R}^3$ is a linear transformation given by

$$T\vec{x} = A\vec{x}$$

where A is a 3×3 matrix. Show that T is an isomorphism if and only if A is invertible.

Exercise 4.E.56

Suppose $T: \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation given by

$$T\vec{x} = A\vec{x}$$

where A is an $m \times n$ matrix. Show that T is never an ismorphism if $m \neq n$. In particular, show that if $m > n$, T cannot be onto and if $m < n$, then T cannot be one to one.

Exercise 4.E.57

Define $T: \mathbb{R}^2 \to \mathbb{R}^3$ as follows.

$$T\vec{x} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \vec{x}$$

where on the right, it is just matrix multiplication of the vector $\vec{x}$ which is meant. Show that T is one to one. Next let $W = im(T)$. Show that T is an isomorphism of $\mathbb{R}^2$ and $im(T)$.

Exercise 4.E.58

In the above problem, find a 2×3 matrix A such that the restriction of A to $im(T)$ gives the same result as T^{-1} on $im(T)$. Hint: You might let A be such that

$$A \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \; A \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

now find another vector $\vec{v} \in \mathbb{R}^3$ such that

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \vec{v} \right\}$$

is a basis. You could pick

$$\vec{v} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

for example. Explain why this one works or one of your choice works. Then you could define $A\vec{v}$ to equal some vector in $\mathbb{R}^2$. Explain why there will be more than one such matrix A which will deliver the inverse isomorphism T^{-1} on $im(T)$.

Exercise 4.E.59

Now let V equan $span\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right\}$ and let T: $V \to W$ be a linear transformation where

$$W = span\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}\right\}$$

and

$$T\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, T\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Explain why T is an isomorphism. Determine a matrix A which, when multiplied on the left gives the same result as T on V and a matrix B which delivers T^{-1} on W. Hint: You need to have

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

Now enlarge $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ to obtain a basis for $\mathbb{R}^3$. You could add in $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ for example, and then pick another vector in $\mathbb{R}^4$ and let $A\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ equal this other vector. Then you would have

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

This would involve picking for the new vector in $\mathbb{R}^4$ the vector $\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T$. Then you could find A. You can do something similar to find a matrix for T^{-1} denoted as B.

Exercise 4.E.60

Let $V = \mathbb{R}^3$ and let

$$W = span(S), \text{ where } S = \left\{\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}\right\}$$

Find a basis of W consisting of vectors in S.

Answer

In this case $\dim(W) = 1$ and a basis for W consisting of vectors in S can be obtained by taking any (nonzero) vector from S.

Exercise 4.E.61

Let T be a linear transformation given by

$$T\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

Find a basis for $\ker(T)$ and $im(T)$.

Answer

A basis for $\ker(T)$ is $\left\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\}$ and a basis for $im(T)$ is $\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$. There are many other possibilities for the specific bases, but in this case $\dim(\ker(T)) = 1$ and $\dim(im(T)) = 1$.

Exercise 4.E.62

Let T be a linear transformation given by

$$T\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

Find a basis for $\ker(T)$ and $im(T)$.

Answer

In this case $\ker(T) = \{0\}$ and $im(T) = \mathbb{R}^2$ (pick any basis of $\mathbb{R}^2$).

Exercise 4.E.63

Let $V = \mathbb{R}^3$ and let

$$W = span\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}\right\}$$

Extend this basis of W to a basis of V.

Answer

There are many possible such extensions, one is (how do we know?):

$$\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}$$

Exercise 4.E.64

Let T be a linear transformation given by

https://math.libretexts.org/@go/page/170465

$$T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

What is $\dim(\ker(T))$?

Answer

We can easily see that $\dim(im(T)) = 1$, and thus $\dim(\ker(T)) = 3 - \dim(im(T)) = 3 - 1 = 2$.

Exercise 4.E.65

Let $B = \left\{ \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix} \right\}$ be a basis of $\mathbb{R}^2$ and let $\vec{x} = \begin{bmatrix} 5 \\ -7 \end{bmatrix}$ be a vector in $\mathbb{R}^2$. Find $C_B(\vec{x})$.

Exercise 4.E.66

Let $B = \left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \right\}$ be a basis of $\mathbb{R}^3$ and let $\vec{x} = \begin{bmatrix} 5 \\ -1 \\ 4 \end{bmatrix}$ be a vector in $\mathbb{R}^2$. Find $C_B(\vec{x})$.

Answer

$$C_B(\vec{x}) = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

Exercise 4.E.67

Let T: $\mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation defined by $T\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} a+b \\ a-b \end{bmatrix}$.

Consider the two bases

$$B_1 = \{\vec{v}_1, \vec{v}_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$$

and

$$B_2 = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$$

Find the matrix M_{B_2,B_1} of T with respect to the bases B_1 and B_2.

Answer

$$M_{B_2B_1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$$

Exercise 4.E.68

Write the solution set of the following system as a linear combination of vectors

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

Answer

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

Exercise 4.E.69

Using Exercise 4.E.68 find the general solution to the following linear system.

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

Answer

Note that this has the same matrix as the above problem. Solution is: $\begin{bmatrix} -3\hat{t}_3 \\ -\hat{t}_3 \\ \hat{t}_3 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$, $\hat{t}_3 \in \mathbb{R}$

Exercise 4.E.70

Write the solution set of the following system as a linear combination of vectors

$$\begin{bmatrix} 0 & -1 & 2 \\ 1 & -2 & 1 \\ 1 & -4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Answer

Solution is: $\begin{bmatrix} 3\hat{t} \\ 2\hat{t} \\ \hat{t} \end{bmatrix}$, A basis is $\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$

Exercise 4.E.71

Using Exercise 4.E.70 find the general solution to the following linear system.

$$\begin{bmatrix} 0 & -1 & 2 \\ 1 & -2 & 1 \\ 1 & -4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

Answer

Solution is: $\begin{bmatrix} 3\hat{t} \\ 2\hat{t} \\ \hat{t} \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ 0 \end{bmatrix}$, $\hat{t} \in \mathbb{R}$

Exercise 4.E.72

Write the solution set of the following system as a linear combination of vectors.

$$\begin{bmatrix} 1 & -1 & 2 \\ 1 & -2 & 0 \\ 3 & -4 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Answer

Solution is: $\begin{bmatrix} -4\hat{t} \\ -2\hat{t} \\ \hat{t} \end{bmatrix}$. A basis is $\begin{bmatrix} -4 \\ -2 \\ 1 \end{bmatrix}$

? Exercise 4.E.73

Using Exercise 4.E.72 find the general solution to the following linear system.

$$\begin{bmatrix} 1 & -1 & 2 \\ 1 & -2 & 0 \\ 3 & -4 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

Answer

Solution is: $\begin{bmatrix} -4\hat{t} \\ -2\hat{t} \\ \hat{t} \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \ \hat{t} \in \mathbb{R}$

? Exercise 4.E.74

Write the solution set of the following system as a linear combination of vectors

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

Answer

Solution is: $\begin{bmatrix} -\hat{t} \\ 2\hat{t} \\ \hat{t} \end{bmatrix}, \ \hat{t} \in \mathbb{R}$

? Exercise 4.E.75

Using Exercise 4.E.74 find the general solution to the following linear system.

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

Answer

Solution is: $\begin{bmatrix} -\hat{t} \\ 2\hat{t} \\ \hat{t} \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$

? Exercise 4.E.76

Write the solution set of the following system as a linear combination of vectors

https://math.libretexts.org/@go/page/170465

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

Answer

Solution is: $\begin{bmatrix} 0 \\ -\hat{t} \\ -\hat{t} \\ \hat{t} \end{bmatrix}$, $\hat{t} \in \mathbb{R}$

Exercise 4.E.77

Using Exercise 4.E.76 find the general solution to the following linear system.

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

Answer

Solution is: $\begin{bmatrix} 0 \\ -\hat{t} \\ -\hat{t} \\ \hat{t} \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \\ -1 \\ 0 \end{bmatrix}$

Exercise 4.E.78

Write the solution set of the following system as a linear combination of vectors

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

Answer

Solution is: $\begin{bmatrix} -s-t \\ s \\ s \\ t \end{bmatrix}$, $s, t \in \mathbb{R}$. A basis is

$$\left\{ \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Exercise 4.E.79

Using Exercise 4.E.78 find the general solution to the following linear system.

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

Answer

Solution is:

$$\begin{bmatrix} -\hat{t} \\ \hat{t} \\ \hat{t} \\ 0 \end{bmatrix} + \begin{bmatrix} -8 \\ 5 \\ 0 \\ 5 \end{bmatrix}$$

Exercise $4.E.80$

Write the solution set of the following system as a linear combination of vectors

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

Answer

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

$$\left\{ \begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Exercise $4.E.81$

Using Exercise $4.E.80$ find the general solution to the following linear system.

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

Answer

Solution is:

https://math.libretexts.org/@go/page/170465

$$\begin{bmatrix} \frac{3}{2} \\ -\frac{1}{2} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -\frac{1}{2}s - \frac{1}{2}t \\ \frac{1}{2}s - \frac{1}{2}t \\ s \\ t \end{bmatrix}$$

Exercise 4.E.82

Write the solution set of the following system as a linear combination of vectors

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 2 & 1 & 1 & 2 \\ 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Answer

Solution is: $\begin{bmatrix} -\hat{t} \\ \hat{t} \\ \hat{t} \\ 0 \end{bmatrix}$, a basis is $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$.

Exercise 4.E.83

Using Exercise 4.E.82 find the general solution to the following linear system.

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

Answer

Solution is: $\begin{bmatrix} -\hat{t} \\ \hat{t} \\ \hat{t} \\ 0 \end{bmatrix} + \begin{bmatrix} -9 \\ 5 \\ 0 \\ 6 \end{bmatrix}, t \in \mathbb{R}$.

Exercise 4.E.84

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.

Answer

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 $A\vec{x} = \vec{b}$.

This page titled 4.E: Exercises is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) .

CHAPTER OVERVIEW

5: Vector Spaces and Subspaces

5.1: Vector Spaces
5.2: Subspaces
5.3: Basis and Dimension
5.4: Rank and Homogeneous Systems
5.5: The Rank Theorem
5.6: Bases as Coordinate Systems
5.7: Spanning, Linear Independence and Basis in R^n
5.8: A Linear Transformations
5.9: A Isomorphisms
5.10: The Kernel and Image of a Linear Map
5.11: The Matrix of a Linear Transformation
5.E: Exercises

This page titled 5: Vector Spaces and Subspaces is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

5.1: Vector Spaces

Outcomes

A. Develop the abstract concept of a vector space through axioms.
B. Deduce basic properties of vector spaces.
C. Use the vector space axioms to determine if a set and its operations constitute a vector space.

In this section we consider the idea of an abstract vector space. A vector space is something which has two operations satisfying the following vector space axioms.

Definition 5.1.1: Vector Space

A vector space V is a set of vectors with two operations defined, addition and scalar multiplication, which satisfy the axioms of addition and scalar multiplication.

In the following definition we define two operations; vector addition, denoted by $+$ and scalar multiplication denoted by placing the scalar next to the vector. A vector space need not have usual operations, and for this reason the operations will always be given in the definition of the vector space. The below axioms for addition (written +) and scalar multiplication must hold for however addition and scalar multiplication are defined for the vector space.

It is important to note that we have seen much of this content before, in terms of $\mathbb{R}^n$. We will prove in this section that $\mathbb{R}^n$ is an example of a vector space and therefore all discussions in this chapter will pertain to $\mathbb{R}^n$. While it may be useful to consider all concepts of this chapter in terms of $\mathbb{R}^n$, it is also important to understand that these concepts apply to *all* vector spaces.

In the following definition, we will choose scalars a, b to be real numbers and are thus dealing with *real* vector spaces. However, we could also choose scalars which are complex numbers. In this case, we would call the vector space V complex.

Definition 5.1.2: Axioms of Addition

Let $\vec{v}, \vec{w}, \vec{z}$ be vectors in a vector space V. Then they satisfy the following axioms of addition:

- Closed under Addition

$$\text{If } \vec{v}, \vec{w} \text{ are in } V, \text{ then } \vec{v}+\vec{w} \text{ is also in } V.$$

- The Commutative Law of Addition

$$\vec{v}+\vec{w}=\vec{w}+\vec{v}$$

- The Associative Law of Addition

$$(\vec{v}+\vec{w})+\vec{z}=\vec{v}+(\vec{w}+\vec{z})$$

- The Existence of an Additive Identity

$$\vec{v}+\vec{0}=\vec{v}$$

- The Existence of an Additive Inverse

$$\vec{v}+(-\vec{v})=\vec{0}$$

vectorspaceaxiomsaddition

Definition 5.1.3: Axioms of Scalar Multiplication

Let $a, b \in \mathbb{R}$ and let $\vec{v}, \vec{w}, \vec{z}$ be vectors in a vector space V. Then they satisfy the following axioms of scalar multiplication:

- Closed under Scalar Multiplication

If a is a real number, and $\vec{v}$ is in V, then $a\vec{v}$ is in V.

- $a\left(\vec{v}+\vec{w}\right)=a\vec{v}+a\vec{w}$
- $\left(a+b\right)\vec{v}=a\vec{v}+b\vec{v}$
- $a\left(b\vec{v}\right)=\left(ab\right)\vec{v}$
- $1\vec{v}=\vec{v}$

Consider the following example, in which we prove that $\mathbb{R}^n$ is in fact a vector space.

Example 5.1.1: $\mathbb{R}^n$

$\mathbb{R}^n$, under the usual operations of vector addition and scalar multiplication, is a vector space.

Solution

To show that $\mathbb{R}^n$ is a vector space, we need to show that the above axioms hold. Let $\vec{x},\vec{y},\vec{z}$ be vectors in $\mathbb{R}^n$. We first prove the axioms for vector addition.

- To show that $\mathbb{R}^n$ is closed under addition, we must show that for two vectors in $\mathbb{R}^n$ their sum is also in $\mathbb{R}^n$. The sum $\vec{x}+\vec{y}$ is given by:

$$\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}+\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}=\begin{bmatrix}x_1+y_1\\x_2+y_2\\\vdots\\x_n+y_n\end{bmatrix}$$

  The sum is a vector with n entries, showing that it is in $\mathbb{R}^n$. Hence $\mathbb{R}^n$ is closed under vector addition.
- To show that addition is commutative, consider the following:

$$\begin{aligned}\vec{x}+\vec{y}&=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}+\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}\\&=\begin{bmatrix}x_1+y_1\\x_2+y_2\\\vdots\\x_n+y_n\end{bmatrix}\\&=\begin{bmatrix}y_1+x_1\\y_2+x_2\\\vdots\\y_n+x_n\end{bmatrix}\\&=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}+\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}\\&=\vec{y}+\vec{x}\end{aligned}$$

  Hence addition of vectors in $\mathbb{R}^n$ is commutative.
- We will show that addition of vectors in $\mathbb{R}^n$ is associative in a similar way.

$$
\begin{aligned}
(\vec{x}+\vec{y})+\vec{z} &= \left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}\right) + \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \\
&= \begin{bmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \\
&= \begin{bmatrix} (x_1+y_1)+z_1 \\ (x_2+y_2)+z_2 \\ \vdots \\ (x_n+y_n)+z_n \end{bmatrix} \\
&= \begin{bmatrix} x_1+(y_1+z_1) \\ x_2+(y_2+z_2) \\ \vdots \\ x_n+(y_n+z_n) \end{bmatrix} \\
&= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1+z_1 \\ y_2+z_2 \\ \vdots \\ y_n+z_n \end{bmatrix} \\
&= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}\right) \\
&= \vec{x}+(\vec{y}+\vec{z})
\end{aligned}
$$

Hence addition of vectors is associative.

- Next, we show the existence of an additive identity. Let $\vec{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

$$
\begin{aligned}
\vec{x}+\vec{0} &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} x_1+0 \\ x_2+0 \\ \vdots \\ x_n+0 \end{bmatrix} \\
&= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= \vec{x}
\end{aligned}
$$

Hence the zero vector $\vec{0}$ is an additive identity.

- Next, we prove the existence of an additive inverse. Let $-\vec{x} = \begin{pmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{pmatrix}$.

$$\begin{aligned} \vec{x} + (-\vec{x}) &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 - x_1 \\ x_2 - x_2 \\ \vdots \\ x_n - x_n \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= \vec{0} \end{aligned}$$

  Hence $-\vec{x}$ is an additive inverse.

We now need to prove the axioms related to scalar multiplication. Let a, b be real numbers and let $\vec{x}, \vec{y}$ be vectors in $\mathbb{R}^n$.

- We first show that $\mathbb{R}^n$ is closed under scalar multiplication. To do so, we show that $a\vec{x}$ is also a vector with n entries.

$$a\vec{x} = a \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix}$$

  The vector $a\vec{x}$ is again a vector with n entries, showing that $\mathbb{R}^n$ is closed under scalar multiplication.
- We wish to show that $a(\vec{x} + \vec{y}) = a\vec{x} + a\vec{y}$.

$$
\begin{aligned}
a(\vec{x}+\vec{y}) &= a\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \\ \vdots x_n \end{bmatrix}\right) \\
&= a\begin{bmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{bmatrix} \\
&= \begin{bmatrix} a(x_1+y_1) \\ a(x_2+y_2) \\ \vdots \\ a(x_n+y_n) \end{bmatrix} \\
&= \begin{bmatrix} ax_1+ay_1 \\ ax_2+ay_2 \\ \vdots \\ ax_n+ay_n \end{bmatrix} \\
&= \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix} + \begin{bmatrix} ay_1 \\ ay_2 \\ \vdots \\ ay_n \end{bmatrix} \\
&= a\vec{x}+a\vec{y}
\end{aligned}
$$

- Next, we wish to show that $(a+b)\vec{x} = a\vec{x}+b\vec{x}$.

$$
\begin{aligned}
(a+b)\vec{x} &= (a+b)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} (a+b)x_1 \\ (a+b)x_2 \\ \vdots \\ (a+b)x_n \end{bmatrix} \\
&= \begin{bmatrix} ax_1+bx_1 \\ ax_2+bx_2 \\ \vdots \\ ax_n+bx_n \end{bmatrix} \\
&= \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix} + \begin{bmatrix} bx_1 \\ bx_2 \\ \vdots \\ bx_n \end{bmatrix} \\
&= a\vec{x}+b\vec{x}
\end{aligned}
$$

- We wish to show that $a(b\vec{x}) = (ab)\vec{x}$.

$$
\begin{aligned}
a(b\vec{x}) &= a\left(b\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) \\
&= a\left(\begin{bmatrix} bx_1 \\ bx_2 \\ \vdots \\ bx_n \end{bmatrix}\right) \\
&= \begin{bmatrix} a(bx_1) \\ a(bx_2) \\ \vdots \\ a(bx_n) \end{bmatrix} \\
&= \begin{bmatrix} (ab)x_1 \\ (ab)x_2 \\ \vdots \\ (ab)x_n \end{bmatrix} \\
&= (ab)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= (ab)\vec{x}
\end{aligned}
$$

- Finally, we need to show that $1\vec{x} = \vec{x}$.

$$
\begin{aligned}
1\vec{x} &= 1\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} 1x_1 \\ 1x_2 \\ \vdots \\ 1x_n \end{bmatrix} \\
&= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= \vec{x}
\end{aligned}
$$

By the above proofs, it is clear that $\mathbb{R}^n$ satisfies the vector space axioms. Hence, $\mathbb{R}^n$ is a vector space under the usual operations of vector addition and scalar multiplication.

We now consider some examples of vector spaces.

Example 5.1.2: Vector Space of Polynomials

Let $\mathbb{P}_2$ be the set of all polynomials of at most degree 2 as well as the zero polynomial. Define addition to be the standard addition of polynomials, and scalar multiplication the usual multiplication of a polynomial by a number. Then $\mathbb{P}_2$ is a vector space.

Solution

We can write $\mathbb{P}_2$ explicitly as

$$\mathbb{P}_2 = \left\{a_2x^2 + a_1x + a_0 | a_i \in \mathbb{R} \text{ for all } i\right\}$$

To show that $\mathbb{P}_2$ is a vector space, we verify the axioms. Let $p(x), q(x), r(x)$ be polynomials in $\mathbb{P}_2$ and let a, b, c be real numbers. Write $p(x) = p_2x^2 + p_1x + p_0$, $q(x) = q_2x^2 + q_1x + q_0$, and $r(x) = r_2x^2 + r_1x + r_0$.

- We first prove that addition of polynomials in $\mathbb{P}_2$ is closed. For two polynomials in $\mathbb{P}_2$ we need to show that their sum is also a polynomial in $\mathbb{P}_2$. From the definition of $\mathbb{P}_2$, a polynomial is contained in $\mathbb{P}_2$ if it is of degree at most 2 or the zero polynomial.

$$\begin{aligned} p(x) + q(x) &= p_2x^2 + p_1x + p_0 + q_2x^2 + q_1x + q_0 \\ &= (p_2 + q_2)x^2 + (p_1 + q_1)x + (p_0 + q_0) \end{aligned}$$

  The sum is a polynomial of degree 2 and therefore is in $\mathbb{P}_2$. It follows that $\mathbb{P}_2$ is closed under addition.
- We need to show that addition is commutative, that is $p(x) + q(x) = q(x) + p(x)$.

$$\begin{aligned} p(x) + q(x) &= p_2x^2 + p_1x + p_0 + q_2x^2 + q_1x + q_0 \\ &= (p_2 + q_2)x^2 + (p_1 + q_1)x + (p_0 + q_0) \\ &= (q_2 + p_2)x^2 + (q_1 + p_1)x + (q_0 + p_0) \\ &= q_2x^2 + q_1x + q_0 + p_2x^2 + p_1x + p_0 \\ &= q(x) + p(x) \end{aligned}$$

- Next, we need to show that addition is associative. That is, that $(p(x) + q(x)) + r(x) = p(x) + (q(x) + r(x))$.

$$\begin{aligned} (p(x) + q(x)) + r(x) &= \left(p_2x^2 + p_1x + p_0 + q_2x^2 + q_1x + q_0\right) + r_2x^2 + r_1x + r_0 \\ &= (p_2 + q_2)x^2 + (p_1 + q_1)x + (p_0 + q_0) + r_2x^2 + r_1x + r_0 \\ &= (p_2 + q_2 + r_2)x^2 + (p_1 + q_1 + r_1)x + (p_0 + q_0 + r_0) \\ &= p_2x^2 + p_1x + p_0 + (q_2 + r_2)x^2 + (q_1 + r_1)x + (q_0 + r_0) \\ &= p_2x^2 + p_1x + p_0 + \left(q_2x^2 + q_1x + q_0 + r_2x^2 + r_1x + r_0\right) \\ &= p(x) + (q(x) + r(x)) \end{aligned}$$

- Next, we must prove that there exists an additive identity. Let $0(x) = 0x^2 + 0x + 0$.

$$\begin{aligned} p(x) + 0(x) &= p_2x^2 + p_1x + p_0 + 0x^2 + 0x + 0 \\ &= (p_2 + 0)x^2 + (p_1 + 0)x + (p_0 + 0) \\ &= p_2x^2 + p_1x + p_0 \\ &= p(x) \end{aligned}$$

  Hence an additive identity exists, specifically the zero polynomial.
- Next we must prove that there exists an additive inverse. Let $-p(x) = -p_2x^2 - p_1x - p_0$ and consider the following:

$$\begin{aligned} p(x) + (-p(x)) &= p_2x^2 + p_1x + p_0 + \left(-p_2x^2 - p_1x - p_0\right) \\ &= (p_2 - p_2)x^2 + (p_1 - p_1)x + (p_0 - p_0) \\ &= 0x^2 + 0x + 0 \\ &= 0(x) \end{aligned}$$

  Hence an additive inverse $-p(x)$ exists such that $p(x) + (-p(x)) = 0(x)$.

We now need to verify the axioms related to scalar multiplication.

- First we prove that $\mathbb{P}_2$ is closed under scalar multiplication. That is, we show that $ap(x)$ is also a polynomial of degree at most 2.

$$ap(x) = a\left(p_2x^2 + p_1x + p_0\right) = ap_2x^2 + ap_1x + ap_0$$

  Therefore $\mathbb{P}_2$ is closed under scalar multiplication.
- We need to show that $a(p(x) + q(x)) = ap(x) + aq(x)$.

$$\begin{aligned} a(p(x)+q(x)) &= a\left(p_2x^2+p_1x+p_0+q_2x^2+q_1x+q_0\right) \\ &= a\left((p_2+q_2)x^2+(p_1+q_1)x+(p_0+q_0)\right) \\ &= a(p_2+q_2)x^2+a(p_1+q_1)x+a(p_0+q_0) \\ &= (ap_2+aq_2)x^2+(ap_1+aq_1)x+(ap_0+aq_0) \\ &= ap_2x^2+ap_1x+ap_0+aq_2x^2+aq_1x+aq_0 \\ &= ap(x)+aq(x) \end{aligned}$$

- Next we show that $(a+b)p(x)=ap(x)+bp(x)$.

$$\begin{aligned} (a+b)p(x) &= (a+b)(p_2x^2+p_1x+p_0) \\ &= (a+b)p_2x^2+(a+b)p_1x+(a+b)p_0 \\ &= ap_2x^2+ap_1x+ap_0+bp_2x^2+bp_1x+bp_0 \\ &= ap(x)+bp(x) \end{aligned}$$

- The next axiom which needs to be verified is $a(bp(x))=(ab)p(x)$.

$$\begin{aligned} a(bp(x)) &= a\left(b\left(p_2x^2+p_1x+p_0\right)\right) \\ &= a\left(bp_2x^2+bp_1x+bp_0\right) \\ &= abp_2x^2+abp_1x+abp_0 \\ &= (ab)(p_2x^2+p_1x+p_0) \\ &= (ab)p(x) \end{aligned}$$

- Finally, we show that $1p(x)=p(x)$.

$$\begin{aligned} 1p(x) &= 1\left(p_2x^2+p_1x+p_0\right) \\ &= 1p_2x^2+1p_1x+1p_0 \\ &= p_2x^2+p_1x+p_0 \\ &= p(x) \end{aligned}$$

Since the above axioms hold, we know that $\mathbb{P}_2$ as described above is a vector space.

Another important example of a vector space is the set of all matrices of the same size.

Example 5.1.3: Vector Space of Matrices

Let $\mathbb{M}_{2,3}$ be the set of all 2×3 matrices. Using the usual operations of matrix addition and scalar multiplication, show that $\mathbb{M}_{2,3}$ is a vector space.

Solution

Let A,B be 2×3 matrices in $\mathbb{M}_{2,3}$. We first prove the axioms for addition.

- In order to prove that $\mathbb{M}_{2,3}$ is closed under matrix addition, we show that the sum $A+B$ is in $\mathbb{M}_{2,3}$. This means showing that $A+B$ is a 2×3 matrix.

$$\begin{aligned} A+B &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \end{bmatrix} \end{aligned}$$

  You can see that the sum is a 2×3 matrix, so it is in $\mathbb{M}_{2,3}$. It follows that $\mathbb{M}_{2,3}$ is closed under matrix addition.
- The remaining axioms regarding matrix addition follow from properties of matrix addition. Therefore $\mathbb{M}_{2,3}$ satisfies the axioms of matrix addition.

We now turn our attention to the axioms regarding scalar multiplication. Let A,B be matrices in $\mathbb{M}_{2,3}$ and let c be a real number.

- We first show that $\mathbb{M}_{2,3}$ is closed under scalar multiplication. That is, we show that cA a 2×3 matrix.

$$cA = c\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = \begin{bmatrix} ca_{11} & ca_{12} & ca_{13} \\ ca_{21} & ca_{22} & ca_{23} \end{bmatrix}$$

This is a 2×3 matrix in $\mathbb{M}_{2,3}$ which proves that the set is closed under scalar multiplication.

- The remaining axioms of scalar multiplication follow from properties of scalar multiplication of matrices. Therefore $\mathbb{M}_{2,3}$ satisfies the axioms of scalar multiplication.

In conclusion, $\mathbb{M}_{2,3}$ satisfies the required axioms and is a vector space.

While here we proved that the set of all 2×3 matrices is a vector space, there is nothing special about this choice of matrix size. In fact if we instead consider $\mathbb{M}_{m,n}$, the set of all $m \times n$ matrices, then $\mathbb{M}_{m,n}$ is a vector space under the operations of matrix addition and scalar multiplication.

We now examine an example of a set that does not satisfy all of the above axioms, and is therefore *not* a vector space.

Example 5.1.4: Not a Vector Space

Let V denote the set of 2×3 matrices. Let addition in V be defined by $A + B = A$ for matrices A, B in V. Let scalar multiplication in V be the usual scalar multiplication of matrices. Show that V is not a vector space.

Solution

In order to show that V is not a vector space, it suffices to find only one axiom which is not satisfied. We will begin by examining the axioms for addition until one is found which does not hold. Let A, B be matrices in V.

- We first want to check if addition is closed. Consider $A + B$. By the definition of addition in the example, we have that $A + B = A$. Since A is a 2×3 matrix, it follows that the sum $A + B$ is in V, and V is closed under addition.
- We now wish to check if addition is commutative. That is, we want to check if $A + B = B + A$ **for all** choices of A and B in V. From the definition of addition, we have that $A + B = A$ and $B + A = B$. Therefore, we can find A, B in V such that these sums are not equal. One example is

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Using the operation defined by $A + B = A$, we have

$$\begin{aligned} A + B &= A \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ B + A &= B \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{aligned}$$

It follows that $A + B \neq B + A$. Therefore addition as defined for V is not commutative and V fails this axiom. Hence V is not a vector space.

Consider another example of a vector space.

Example 5.1.5: Vector Space of Functions

Let S be a nonempty set and define $\mathbb{F}_S$ to be the set of *real* functions defined on S. In other words, we write $\mathbb{F}_S : S \mapsto \mathbb{R}$. Letting a, b, c be scalars and f, g, h functions, the vector operations are defined as

$$\begin{aligned} (f + g)(x) &= f(x) + g(x) \\ (af)(x) &= a(f(x)) \end{aligned}$$

Show that $\mathbb{F}_S$ is a vector space.

Solution

To verify that $\mathbb{F}_S$ is a vector space, we must prove the axioms beginning with those for addition. Let f, g, h be functions in $\mathbb{F}_S$.

- First we check that addition is closed. For functions f, g defined on the set S, their sum given by
$$(f+g)(x) = f(x)+g(x)$$
is again a function defined on S. Hence this sum is in $\mathbb{F}_S$ and $\mathbb{F}_S$ is closed under addition.
- Secondly, we check the commutative law of addition:
$$(f+g)(x) = f(x)+g(x) = g(x)+f(x) = (g+f)(x)$$
Since x is arbitrary, $f+g=g+f$.
- Next we check the associative law of addition:
$$((f+g)+h)(x) = (f+g)(x)+h(x) = (f(x)+g(x))+h(x)$$
$$= f(x)+(g(x)+h(x)) = (f(x)+(g+h)(x)) = (f+(g+h))(x)$$
and so $(f+g)+h = f+(g+h)$.
- Next we check for an additive identity. Let 0 denote the function which is given by $0(x) = 0$. Then this is an additive identity because
$$(f+0)(x) = f(x)+0(x) = f(x)$$
and so $f+0 = f$.
- Finally, check for an additive inverse. Let $-f$ be the function which satisfies $(-f)(x) = -f(x)$. Then
$$(f+(-f))(x) = f(x)+(-f)(x) = f(x)+-f(x) = 0$$
Hence $f+(-f) = 0$.

Now, check the axioms for scalar multiplication.

- We first need to check that $\mathbb{F}_S$ is closed under scalar multiplication. For a function $f(x)$ in $\mathbb{F}_S$ and real number a, the function $(af)(x) = a(f(x))$ is again a function defined on the set S. Hence $a(f(x))$ is in $\mathbb{F}_S$ and $\mathbb{F}_S$ is closed under scalar multiplication.
- $$((a+b)f)(x) = (a+b)f(x) = af(x)+bf(x) = (af+bf)(x)$$
and so $(a+b)f = af+bf$.
- $$(a(f+g))(x) = a(f+g)(x) = a(f(x)+g(x))$$
$$= af(x)+bg(x) = (af+bg)(x)$$
and so $a(f+g) = af+bg$.
- $$((ab)f)(x) = (ab)f(x) = a(bf(x)) = (a(bf))(x)$$
so $(abf) = a(bf)$.
- Finally $(1f)(x) = 1f(x) = f(x)$ so $1f = f$.

It follows that V satisfies all the required axioms and is a vector space.

Consider the following important theorem.

Theorem 5.1.1: Uniqueness

In any vector space, the following are true:

1. $\vec{0}$, the additive identity, is unique
2. $-\vec{x}$, the additive inverse, is unique
3. $0\vec{x} = \vec{0}$ for all vectors $\vec{x}$

4. $(-1)\,\vec{x} = -\vec{x}$ for all vectors $\vec{x}$

Proof

1. When we say that the additive identity, $\vec{0}$, is unique, we mean that if a vector acts like the additive identity, then it **is** the additive identity. To prove this uniqueness, we want to show that another vector which acts like the additive identity is actually equal to $\vec{0}$. Suppose $\vec{0}'$ is also an additive identity. Then,

$$\vec{0} + \vec{0}' = \vec{0}$$

Now, for $\vec{0}$ the additive identity given above in the axioms, we have that

$$\vec{0}' + \vec{0} = \vec{0}'$$

So by the commutative property:

$$0 = 0 + 0' = 0' + 0 = 0'$$

This says that if a vector acts like an additive identity (such as $\vec{0}'$), it in fact equals $\vec{0}$. This proves the uniqueness of $\vec{0}$.

2. When we say that the additive inverse, $-\vec{x}$, is unique, we mean that if a vector acts like the additive inverse, then it **is** the additive inverse. Suppose that $\vec{y}$ acts like an additive inverse:

$$\vec{x} + \vec{y} = \vec{0}$$

Then the following holds:

$$\vec{y} = \vec{0} + \vec{y} = (-\vec{x} + \vec{x}) + \vec{y} = -\vec{x} + (\vec{x} + \vec{y}) = -\vec{x} + \vec{0} = -\vec{x}$$

Thus if $\vec{y}$ acts like the additive inverse, it is equal to the additive inverse $-\vec{x}$. This proves the uniqueness of $-\vec{x}$.

3. This statement claims that for all vectors $\vec{x}$, scalar multiplication by 0 equals the zero vector $\vec{0}$. Consider the following, using the fact that we can write $0 = 0 + 0$:

$$0\vec{x} = (0 + 0)\,\vec{x} = 0\vec{x} + 0\vec{x}$$

We use a small trick here: add $-0\vec{x}$ to both sides. This gives

$$\begin{aligned} 0\vec{x} + (-0\vec{x}) &= 0\vec{x} + 0\vec{x} + (-\vec{x}) \\ \vec{0} + 0 &= 0\vec{x} + 0 \\ \vec{0} &= 0\vec{x} \end{aligned}$$

This proves that scalar multiplication of any vector by 0 results in the zero vector $\vec{0}$.

4. Finally, we wish to show that scalar multiplication of -1 and any vector $\vec{x}$ results in the additive inverse of that vector, $-\vec{x}$. Recall from 2. above that the additive inverse is unique. Consider the following:

$$\begin{aligned} (-1)\,\vec{x} + \vec{x} &= (-1)\,\vec{x} + 1\vec{x} \\ &= (-1 + 1)\,\vec{x} \\ &= 0\vec{x} \\ &= \vec{0} \end{aligned}$$

By the uniqueness of the additive inverse shown earlier, any vector which acts like the additive inverse must be equal to the additive inverse. It follows that $(-1)\,\vec{x} = -\vec{x}$.

An important use of the additive inverse is the following theorem.

Theorem 5.1.2

Let V be a vector space. Then $\vec{v} + \vec{w} = \vec{v} + \vec{z}$ implies that $\vec{w} = \vec{z}$ for all $\vec{v}, \vec{w}, \vec{z} \in V$

Proof

The proof follows from the vector space axioms, in particular the existence of an additive inverse $(-\vec{u})$. The proof is left as an exercise to the reader.

This page titled 5.1: Vector Spaces is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **9.1: Algebraic Considerations** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

5.2: Subspaces

Objectives

1. Learn the definition of a subspace.
2. Learn to determine whether or not a subset is a subspace.
3. Learn the most important examples of subspaces.
4. Learn to write a given subspace as a column space or null space.
5. *Recipe:* compute a spanning set for a null space.
6. *Picture:* whether a subset of $\mathbb{R}^2$ or $\mathbb{R}^3$ is a subspace or not.
7. *Vocabulary words:* ***subspace***, ***column space***, ***null space***.

In this section we discuss *subspaces* of $\mathbb{R}^n$. A subspace turns out to be exactly the same thing as a span, except we don't have a particular set of spanning vectors in mind. This change in perspective is quite useful, as it is easy to produce subspaces that are not obviously spans. For example, the solution set of the equation $x+3y+z=0$ is a span because the equation is homogeneous, but we would have to compute the parametric vector form in order to write it as a span.

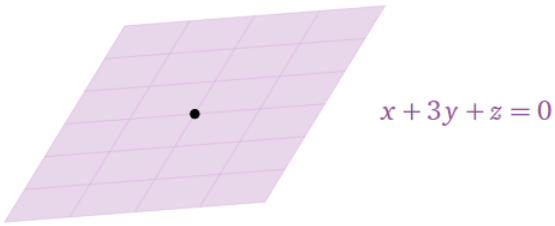


Figure 5.2.1

(A subspace also turns out to be the same thing as the solution set of a homogeneous system of equations.)

Subspaces: Definition and Examples

Definition 5.2.1: Subset

A ***subset*** of $\mathbb{R}^n$ is any collection of points of $\mathbb{R}^n$.

For instance, the unit circle

$$C = \left\{(x,y) \text{ in } \mathbb{R}^2 \;\middle|\; x^2+y^2=1\right\}$$

is a subset of $\mathbb{R}^2$.

Above we expressed C in set builder notation, Note 2.2.3 in Section 2.2: in English, it reads "C is the set of all ordered pairs (x,y) in $\mathbb{R}^2$ such that $x^2+y^2=1$."

Definition 5.2.2: Subspace

A ***subspace*** of $\mathbb{R}^n$ is a subset V of $\mathbb{R}^n$ satisfying:

1. ***Non-emptiness:*** The zero vector is in V.
2. ***Closure under addition:*** If u and v are in V, then $u+v$ is also in V.
3. ***Closure under scalar multiplication:*** If v is in V and c is in $\mathbb{R}$, then cv is also in V.

As a consequence of these properties, we see:

- If v is a vector in V, then all scalar multiples of v are in V by the third property. In other words the line through any nonzero vector in V is also contained in V.

- If u, v are vectors in V and c, d are scalars, then cu, dv are also in V by the third property, so $cu + dv$ is in V by the second property. Therefore, all of $\text{Span}\{u, v\}$ is contained in V
- Similarly, if $v_1, v_2, \ldots, v_n$ are all in V, then $\text{Span}\{v_1, v_2, \ldots, v_n\}$ is contained in V. In other words, *a subspace contains the span of any vectors in it.*

If you choose enough vectors, then eventually their span will fill up V, so we already see that a *subspace is a span*. See this Theorem 5.2.1 below for a precise statement.

Remark

Suppose that V is a non-empty subset of $\mathbb{R}^n$ that satisfies properties 2 and 3. Let v be any vector in V. Then $0v = 0$ is in V by the third property, so V automatically satisfies property 1. It follows that the only subset of $\mathbb{R}^n$ that satisfies properties 2 and 3 but *not* property 1 is the empty subset $\{\}$. This is why we call the first property "non-emptiness".

Example 5.2.1

The set $\mathbb{R}^n$ is a subspace of itself: indeed, it contains zero, and is closed under addition and scalar multiplication.

Example 5.2.2

The set $\{0\}$ containing only the zero vector is a subspace of $\mathbb{R}^n$: it contains zero, and if you add zero to itself or multiply it by a scalar, you always get zero.

Example 5.2.3: A line through the origin

A line L through the origin is a subspace.

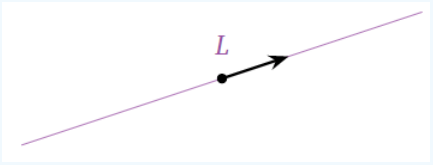


Figure 5.2.2

Indeed, L contains zero, and is easily seen to be closed under addition and scalar multiplication.

Example 5.2.4: A plane through the origin

A plane P through the origin is a subspace.

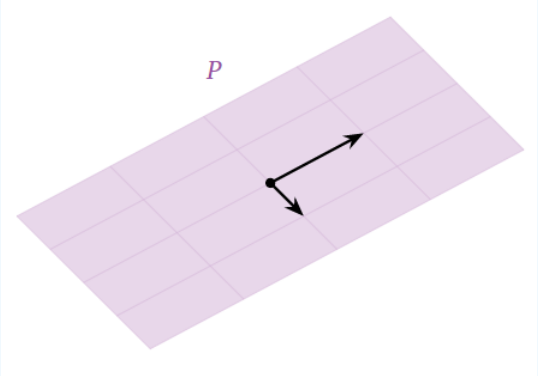


Figure 5.2.3

Indeed, P contains zero; the sum of two vectors in P is also in P; and any scalar multiple of a vector in P is also in P.

Example 5.2.5: Non-example (A line not containing the origin)

A line L (or any other subset) that does not contain the origin is not a subspace. It fails the first defining property: *every subspace contains the origin* by definition.

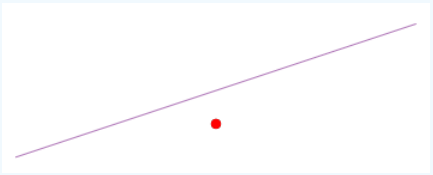


Figure 5.2.4

Example 5.2.6: Non-example (A circle)

The unit circle C is not a subspace. It fails all three defining properties: it does not contain the origin, it is not closed under addition, and it is not closed under scalar multiplication. In the picture, one red vector is the sum of the two black vectors (which are contained in C), and the other is a scalar multiple of a black vector.

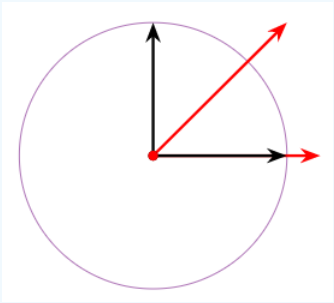


Figure 5.2.5

Example 5.2.7: Non-example (The first quadrant)

The first quadrant in $\mathbb{R}^2$ is not a subspace. It contains the origin and is closed under addition, but it is not closed under scalar multiplication (by negative numbers).

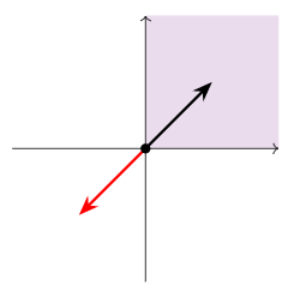


Figure 5.2.6

Example 5.2.8: Non-example (A line union a plane)

The union of a line and a plane in $\mathbb{R}^3$ is not a subspace. It contains the origin and is closed under scalar multiplication, but it is not closed under addition: the sum of a vector on the line and a vector on the plane is not contained in the line or in the plane.

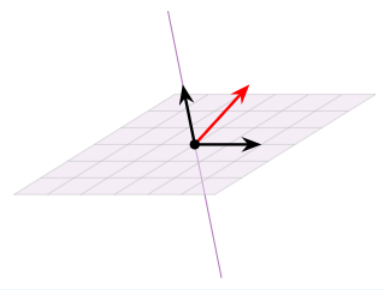


Figure 5.2.7

Note 5.2.1: Subsets versus Subspaces

A subset of $\mathbb{R}^n$ is any collection of vectors whatsoever. For instance, the unit circle

$$C = \left\{(x, y) \text{ in } \mathbb{R}^2 \;\middle|\; x^2 + y^2 = 1\right\}$$

is a subset of $\mathbb{R}^2$, but it is not a subspace. In fact, all of the non-examples above are still subsets of $\mathbb{R}^n$. A subspace is a subset that happens to satisfy the three additional defining properties.

In order to verify that a subset of $\mathbb{R}^n$ is in fact a subspace, one has to check the three defining properties. That is, unless the subset has already been verified to be a subspace: see this important note, Note 5.2.3, below.

Example 5.2.9: Verifying that a subset is a subspace

Let

$$V = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \text{ in } \mathbb{R}^2 | 2a = 3b \right\}.$$

Verify that V is a subspace.

Solution

First we point out that the condition "$2a = 3b$" defines whether or not a vector is in V: that is, to say $\binom{a}{b}$ is in V *means* that $2a = 3b$. In other words, a vector is in V if twice its first coordinate equals three times its second coordinate. So for instance, $\binom{3}{2}$ and $\binom{1/2}{1/3}$ are in V, but $\binom{2}{3}$ is not because $2 \cdot 2 \neq 3 \cdot 3$.

Let us check the first property. The subset V does contain the zero vector $\binom{0}{0}$, because $2 \cdot 0 = 3 \cdot 0$.

Next we check the second property. To show that V is closed under addition, we have to check that for *any* vectors $u = \binom{a}{b}$ and $v = \binom{c}{d}$ in V, the sum $u + v$ is in V. Since we cannot assume anything else about u and v, we must treat them as *unknowns*.

We have

$$\begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} a+c \\ b+d \end{pmatrix}.$$

To say that $\binom{a+c}{b+d}$ is contained in V means that $2(a+c) = 3(b+d)$, or $2a + 2c = 3b + 3d$. The one thing we are allowed to assume about u and v is that $2a = 3b$ and $2c = 3d$, so we see that $u + v$ is indeed contained in V.

Next we check the third property. To show that V is closed under scalar multiplication, we have to check that for any vector $v = \binom{a}{b}$ in V and any scalar c in $\mathbb{R}$, the product cv is in V. Again, we must treat v and c as unknowns. We have

$$c \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} ca \\ cb \end{pmatrix}.$$

To say that $\binom{ca}{cb}$ is contained in V means that $2(ca) = 3(cb)$, i.e., that $c \cdot 2a = c \cdot 3b$. The one thing we are allowed to assume about v is that $2a = 3b$, so cv is indeed contained in V.

Since V satisfies all three defining properties, it is a subspace. In fact, it is the line through the origin with slope $2/3$.

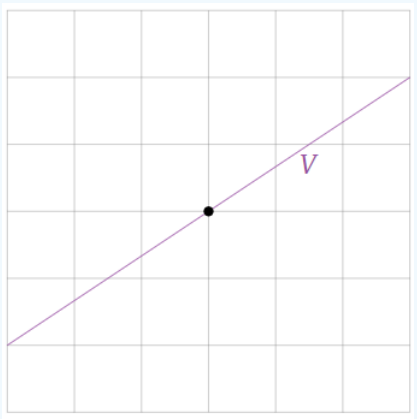


Figure 5.2.8

Example 5.2.10: Showing that a subset is not a subspace

Let

$$V = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \text{ in } \mathbb{R}^2 | ab = 0 \right\}.$$

Is V a subspace?

Solution

First we point out that the condition "$ab = 0$" defines whether or not a vector is in V: that is, to say $\begin{pmatrix} a \\ b \end{pmatrix}$ is in V *means* that $ab = 0$. In other words, a vector is in V if the product of its coordinates is zero, i.e., if one (or both) of its coordinates are zero. So for instance, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$ are in V, but $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is not because $1 \cdot 1 \neq 0$.

Let us check the first property. The subset V does contain the zero vector $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, because $0 \cdot 0 = 0$.

Next we check the third property. To show that V is closed under scalar multiplication, we have to check that for *any* vector $v = \begin{pmatrix} a \\ b \end{pmatrix}$ in V and *any* scalar c in $\mathbb{R}$, the product cv is in V. Since we cannot assume anything else about v and c, we must treat them as *unknowns*.

We have

$$c \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} ca \\ cb \end{pmatrix}.$$

To say that $\begin{pmatrix} ca \\ cb \end{pmatrix}$ is contained in V means that $(ca)(cb) = 0$. Rewriting, this means $c^2(ab) = 0$. The one thing we are allowed to assume about v is that $ab = 0$, so we see that cv is indeed contained in V.

Next we check the second property. It turns out that V is not closed under addition; to verify this, we must show that there exists *some* vectors u, v in V such that $u + v$ is not contained in V. The easiest way to do so is to produce examples of such vectors. We can take $u = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$; these are contained in V because the products of their coordinates are zero, but

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

is not contained in V because $1 \cdot 1 \neq 0$.

Since V does not satisfy the second property (it is not closed under addition), we conclude that V is *not* a subspace. Indeed, it is the union of the two coordinate axes, which is not a span.

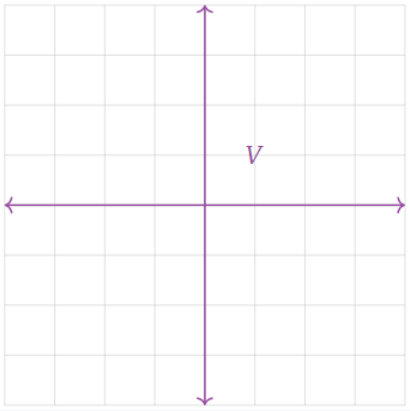


Figure 5.2.9

Common Types of Subspaces

Theorem 5.2.1: Spans are Subspaces and Subspaces are Spans

If $v_1, v_2, \ldots, v_p$ are any vectors in $\mathbb{R}^n$, then $\text{Span}\{v_1, v_2, \ldots, v_p\}$ is a subspace of $\mathbb{R}^n$. Moreover, any subspace of $\mathbb{R}^n$ can be written as a span of a set of p linearly independent vectors in $\mathbb{R}^n$ for $p \leq n$.

Proof

To show that $\text{Span}\{v_1, v_2, \ldots, v_p\}$ is a subspace, we have to verify the three defining properties.

1. The zero vector $0 = 0v_1 + 0v_2 + \cdots + 0v_p$ is in the span.
2. If $u = a_1v_1 + a_2v_2 + \cdots + a_pv_p$ and $v = b_1v_1 + b_2v_2 + \cdots + b_pv_p$ are in $\text{Span}\{v_1, v_2, \ldots, v_p\}$, then
$$u + v = (a_1 + b_1)v_1 + (a_2 + b_2)v_2 + \cdots + (a_p + b_p)v_p$$
is also in $\text{Span}\{v_1, v_2, \ldots, v_p\}$.
3. If $v = a_1v_1 + a_2v_2 + \cdots + a_pv_p$ is in $\text{Span}\{v_1, v_2, \ldots, v_p\}$ and c is a scalar, then
$$cv = ca_1v_1 + ca_2v_2 + \cdots + ca_pv_p$$
is also in $\text{Span}\{v_1, v_2, \ldots, v_p\}$.

Since $\text{Span}\{v_1, v_2, \ldots, v_p\}$ satisfies the three defining properties of a subspace, it is a subspace.

Now let V be a subspace of $\mathbb{R}^n$. If V is the zero subspace, then it is the span of the empty set, so we may assume V is nonzero. Choose a nonzero vector v_1 in V. If $V = \text{Span}\{v_1\}$, then we are done. Otherwise, there exists a vector v_2 that is in V but not in $\text{Span}\{v_1\}$. Then $\text{Span}\{v_1, v_2\}$ is contained in V, and by the increasing span criterion in Section 2.5, Theorem 2.5.2, the set $\{v_1, v_2\}$ is linearly independent. If $V = \text{Span}\{v_1, v_2\}$ then we are done. Otherwise, we continue in this fashion until we have written $V = \text{Span}\{v_1, v_2, \ldots, v_p\}$ for some linearly independent set $\{v_1, v_2, \ldots, v_p\}$. This process terminates after at most n steps by this important note in Section 2.5, Note 2.5.2.

If $V = \text{Span}\{v_1, v_2, \ldots, v_p\}$, we say that V is the subspace ***spanned by*** or ***generated by*** the vectors $v_1, v_2, \ldots, v_p$. We call $\{v_1, v_2, \ldots, v_p\}$ a ***spanning set*** for V.

Any matrix naturally gives rise to *two* subspaces.

Definition 5.2.3: Column Space and Null Space

Let A be an $m \times n$ matrix.

- The ***column space*** of A is the subspace of $\mathbb{R}^m$ spanned by the columns of A. It is written $\text{Col}(A)$.

- The ***null space*** of A is the subspace of $\mathbb{R}^n$ consisting of all solutions of the homogeneous equation $Ax = 0$:

$$\text{Nul}(A) = \{x \text{ in } \mathbb{R}^n \mid Ax = 0\}.$$

The column space is defined to be a span, so it is a subspace by the above Theorem, Theorem 5.2.1. We need to verify that the null space is really a subspace. In Section 2.4 we already saw that the set of solutions of $Ax = 0$ is always a span, so the fact that the null spaces is a subspace should not come as a surprise.

Proof

We have to verify the three defining properties. To say that a vector v is in $\text{Nul}(A)$ *means* that $Av = 0$.

1. The zero vector is in $\text{Nul}(A)$ because $A0 = 0$.
2. Suppose that u, v are in $\text{Nul}(A)$. This means that $Au = 0$ and $Av = 0$. Hence

$$A(u+v) = Au + Av = 0 + 0 = 0$$

by the linearity of the matrix-vector product in Section 2.3, Proposition 2.3.1. Therefore, $u + v$ is in $\text{Nul}(A)$.

3. Suppose that v is in $\text{Nul}(A)$ and c is a scalar. Then

$$A(cv) = cAv = c \cdot 0 = 0$$

by the linearity of the matrix-vector product, Proposition 2.3.1 in Section 2.3, so cv is also in $\text{Nul}(A)$.

Since $\text{Nul}(A)$ satisfies the three defining properties of a subspace, it is a subspace.

Example 5.2.11

Describe the column space and the null space of

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Solution

The column space is the span of the columns of A:

$$\text{Col}(A) = \text{Span}\left\{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right\} = \text{Span}\left\{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right\}.$$

This is a line in $\mathbb{R}^3$.

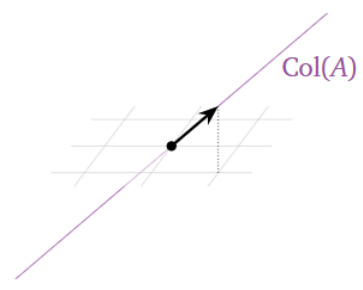


Figure 5.2.10

The null space is the solution set of the homogeneous system $Ax = 0$. To compute this, we need to row reduce A. Its reduced row echelon form is

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

This gives the equation $x + y = 0$, or

$$\begin{cases} x &= -y \\ y &= y \end{cases} \xrightarrow{\text{parametric vector form}} \begin{pmatrix} x \\ y \end{pmatrix} = y \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Hence the null space is $\text{Span}\left\{\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\}$, which is a line in $\mathbb{R}^2$.

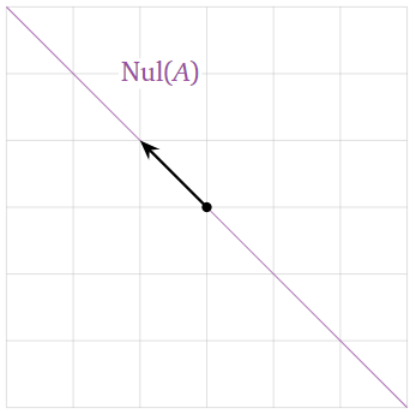


Figure 5.2.11

Notice that the column space is a subspace of $\mathbb{R}^3$, whereas the null space is a subspace of $\mathbb{R}^2$. This is because A has three rows and two columns.

The column space and the null space of a matrix are both subspaces, so they are both spans. The column space of a matrix A is defined to be the span of the columns of A. The null space is defined to be the solution set of $Ax = 0$, so this is a good example of a kind of subspace that we can define *without any spanning set in mind.* In other words, it is easier to show that the null space is a subspace than to show it is a span—see the proof above. In order to do computations, however, it is usually necessary to find a spanning set.

Note 5.2.2: Null Spaces are Solution Sets

The null space of a matrix is the solution set of a homogeneous system of equations. For example, the null space of the matrix

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

is the solution set of $Ax = 0$, i.e., the solution set of the system of equations

$$\begin{cases} x + 7y + 2z = 0 \\ -2x + y + 3z = 0 \\ 4x - 2y - 3z = 0. \end{cases}$$

Conversely, the solution set of any homogeneous system of equations is precisely the null space of the corresponding coefficient matrix.

To find a spanning set for the null space, one has to solve a system of homogeneous equations.

Recipe: Compute a Spanning Set for Null Space

To find a spanning set for $\mathrm{Nul}(A)$, compute the parametric vector form of the solutions to the homogeneous equation $Ax = 0$. The vectors attached to the free variables form a spanning set for $\mathrm{Nul}(A)$.

Example 5.2.12: Two free variables

Find a spanning set for the null space of the matrix

$$A = \begin{pmatrix} 2 & 3 & -8 & -5 \\ -1 & 2 & -3 & -8 \end{pmatrix}.$$

Solution

We compute the parametric vector form of the solutions of $Ax = 0$. The reduced row echelon form of A is

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

The free variables are x_3 and x_4; the parametric form of the solution set is

$$\begin{cases} x_1 = x_3 - 2x_4 \\ x_2 = 2x_3 + 3x_4 \\ x_3 = x_3 \\ x_4 = x_4 \end{cases} \xrightarrow[\text{vector form}]{\text{parametric}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_3 \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -2 \\ 3 \\ 0 \\ 1 \end{pmatrix}.$$

Therefore,

$$\mathrm{Nul}(A) = \mathrm{Span}\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 3 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Example 5.2.13: No free variables

Find a spanning set for the null space of the matrix

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

Solution

We compute the parametric vector form of the solutions of $Ax = 0$. The reduced row echelon form of A is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

There are no free variables; hence the only solution of $Ax = 0$ is the trivial solution. In other words,

$$\mathrm{Nul}(A) = \{0\} = \mathrm{Span}\{0\}.$$

It is natural to define $\mathrm{Span}\{\} = \{0\}$, so that we can take our spanning set to be empty. This is consistent with the definition of dimension, Definition 2.7.2, in Section 2.7.

Writing a subspace as a column space or a null space

A subspace can be given to you in many different forms. In practice, computations involving subspaces are much easier if your subspace is the column space or null space of a matrix. The simplest example of such a computation is finding a spanning set: a column space is by definition the span of the columns of a matrix, and we showed above how to compute a spanning set for a null space using parametric vector form. For this reason, it is useful to rewrite a subspace as a column space or a null space before trying to answer questions about it.

Note 5.2.3

When asking questions about a subspace, it is usually best to rewrite the subspace as a column space or a null space.

This also applies to the question "is my subset a subspace?" If your subset is a column space or null space of a matrix, then the answer is yes.

Example 5.2.14

Let

$$V = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \text{ in } \mathbb{R}^2 | 2a = 3b \right\}$$

be the subset of a previous Example 5.2.9. The subset V is exactly the solution set of the homogeneous equation $2x - 3y = 0$. Therefore,

$$V = \text{Nul}(2 \quad -3)$$

In particular, it is a subspace. The reduced row echelon form of $(2 \quad -3)$ is $(1 \quad -3/2)$, so the parametric form of V is $x = 3/2y$, so the parametric vector form is $\binom{x}{y} = y\binom{3/2}{1}$, and hence $\left\{\binom{3/2}{1}\right\}$ spans V.

Example 5.2.15

Let V be the plane in $\mathbb{R}^3$ defined by

$$V = \left\{ \begin{pmatrix} 2x+y \\ x-y \\ 3x-2y \end{pmatrix} | x, y \text{ are in } \mathbb{R} \right\}.$$

Write V as the column space of a matrix.

Solution

Since

$$\begin{pmatrix} 2x+y \\ x-y \\ 3x-2y \end{pmatrix} = x \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} + y \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix},$$

we notice that V is exactly the span of the vectors

$$\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix}.$$

Hence

$$V = \text{Col} \begin{pmatrix} 2 & 1 \\ 1 & -1 \\ 3 & -2 \end{pmatrix}.$$

This page titled 5.2: Subspaces is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **2.6: Subspaces** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

5.3: Basis and Dimension

Objectives

1. Understand the definition of a basis of a subspace.
2. Understand the basis theorem.
3. *Recipes:* basis for a column space, basis for a null space, basis of a span.
4. *Picture:* basis of a subspace of $\mathbb{R}^2$ or $\mathbb{R}^3$.
5. *Theorem:* basis theorem.
6. *Essential vocabulary words:* ***basis***, ***dimension***.

Basis of a Subspace

As we discussed in Section 2.6, a subspace is the same as a span, except we do not have a set of spanning vectors in mind. There are infinitely many choices of spanning sets for a nonzero subspace; to avoid redundancy, usually it is most convenient to choose a spanning set with the *minimal* number of vectors in it. This is the idea behind the notion of a basis.

Definition 5.3.1: Basis

Let V be a subspace of $\mathbb{R}^n$. A ***basis*** of V is a set of vectors $\{v_1, v_2, \ldots, v_m\}$ in V such that:

1. $V = \text{Span}\{v_1, v_2, \ldots, v_m\}$, and
2. the set $\{v_1, v_2, \ldots, v_m\}$ is linearly independent.

Recall that a set of vectors is *linearly independent* if and only if, when you remove any vector from the set, the span shrinks (Theorem 2.5.1 in Section 2.5). In other words, if $\{v_1, v_2, \ldots, v_m\}$ is a basis of a subspace V, then no proper subset of $\{v_1, v_2, \ldots, v_m\}$ will span V: it is a *minimal* spanning set. Any subspace admits a basis by Theorem 2.6.1 in Section 2.6.

Note 5.3.1

A nonzero subspace has *infinitely many* different bases, but they all contain the same number of vectors.

We leave it as an exercise to prove that any two bases have the same number of vectors; one might want to wait until after learning the invertible matrix theorem in Section 3.5.

Definition 5.3.2: Dimension

Let V be a subspace of $\mathbb{R}^n$. The number of vectors in any basis of V is called the ***dimension*** of V, and is written $\dim V$.

Example 5.3.1: A basis of $\mathbb{R}^2$

Find a basis of $\mathbb{R}^2$.

Solution

We need to find two vectors in $\mathbb{R}^2$ that span $\mathbb{R}^2$ and are linearly independent. One such basis is $\left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}$:

1. They span because any vector $\begin{pmatrix}a\\b\end{pmatrix}$ can be written as a linear combination of $\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}$:
$$\begin{pmatrix}a\\b\end{pmatrix} = a\begin{pmatrix}1\\0\end{pmatrix} + b\begin{pmatrix}0\\1\end{pmatrix}.$$
2. They are linearly independent: if
$$x\begin{pmatrix}1\\0\end{pmatrix} + y\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$$
then $x = y = 0$.

This shows that the plane $\mathbb{R}^2$ has dimension 2.

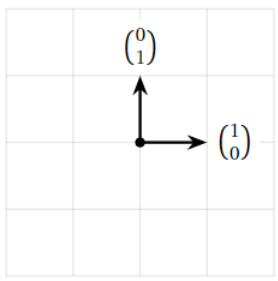


Figure 5.3.1

Example 5.3.2: All bases of $\mathbb{R}^2$

Find all bases of $\mathbb{R}^2$.

Solution

We know from the previous Example 5.3.1 that $\mathbb{R}^2$ has dimension 2, so any basis of $\mathbb{R}^2$ has two vectors in it. Let v_1, v_2 be vectors in $\mathbb{R}^2$, and let A be the matrix with columns v_1, v_2.

1. To say that $\{v_1, v_2\}$ spans $\mathbb{R}^2$ means that A has a pivot, Definition 1.2.5 in Section 1.2, in every *row*: see Theorem 2.3.1 in Section 2.3.
2. To say that $\{v_1, v_2\}$ is linearly independent means that A has a pivot in every *column*: see Recipe: Checking linear independence in Section 2.5.

Since A is a 2×2 matrix, it has a pivot in every row exactly when it has a pivot in every column. Hence any two noncollinear vectors form a basis of $\mathbb{R}^2$. For example,

$$\left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}\right\}$$

is a basis.

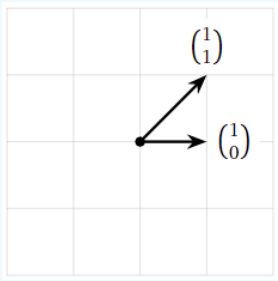


Figure 5.3.2

Example 5.3.3: The standard basis of $\mathbb{R}^n$

One shows exactly as in the above Example 5.3.1 that the standard coordinate vectors

$$e_1 = \begin{pmatrix}1\\0\\\vdots\\0\\0\end{pmatrix}, \quad e_2 = \begin{pmatrix}0\\1\\\vdots\\0\\0\end{pmatrix}, \quad \cdots, \quad e_{n-1} = \begin{pmatrix}0\\0\\\vdots\\1\\0\end{pmatrix}, \quad e_n = \begin{pmatrix}0\\0\\\vdots\\0\\1\end{pmatrix}$$

form a basis for $\mathbb{R}^n$. This is sometimes known as the ***standard basis***.

In particular, $\mathbb{R}^n$ has dimension n.

Example 5.3.4

The previous Example 5.3.3 implies that any basis for $\mathbb{R}^n$ has n vectors in it. Let $v_1, v_2, \ldots, v_n$ be vectors in $\mathbb{R}^n$, and let A be the $n \times n$ matrix with columns $v_1, v_2, \ldots, v_n$.

1. To say that $\{v_1, v_2, \ldots, v_n\}$ spans $\mathbb{R}^n$ means that A has a pivot position, Definition 1.2.5 in Section 1.2, in every *row*: see this Theorem 2.3.1 in Section 2.3.
2. To say that $\{v_1, v_2, \ldots, v_n\}$ is linearly independent means that A has a pivot position in every *column*: see Recipe: Checking linear independence in Section 2.5.

Since A is a square matrix, it has a pivot in every row if and only if it has a pivot in every column. We will see in Section 3.5 that the above two conditions are equivalent to the *invertibility* of the matrix A.

Example 5.3.5

Let

$$V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \text{ in } \mathbb{R}^3 \,|\, x + 3y + z = 0 \right\} \quad \mathcal{B} = \left\{ \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -3 \end{pmatrix} \right\}.$$

Verify that V is a subspace, and show directly that $\mathcal{B}$ is a basis for V.

Solution

First we observe that V is the solution set of the homogeneous equation $x + 3y + z = 0$, so it is a subspace: see this note in Section 2.6, Note 2.6.3. To show that $\mathcal{B}$ is a basis, we really need to verify three things:

0. Both vectors are in V because

$$\begin{aligned} (-3) &+ 3(1) &+ (0) &= 0 \\ (0) &+ 3(1) &+ (-3) &= 0. \end{aligned}$$

1. *Span:* suppose that $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is in V. Since $x + 3y + z = 0$ we have $y = -\frac{1}{3}(x + z)$, so

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ -\frac{1}{3}(x+z) \\ z \end{pmatrix} = -\frac{x}{3}\begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix} - \frac{z}{3}\begin{pmatrix} 0 \\ 1 \\ -3 \end{pmatrix}.$$

   Hence $\mathcal{B}$ spans V.
2. *Linearly independent:*

$$c_1\begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix} + c_2\begin{pmatrix} 0 \\ 1 \\ -3 \end{pmatrix} = 0 \implies \begin{pmatrix} -3c_1 \\ c_1 + c_2 \\ -3c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \implies c_1 = c_2 = 0.$$

   Alternatively, one can observe that the two vectors are not collinear.

Since V has a basis with two vectors, it has dimension two: it is a *plane*.

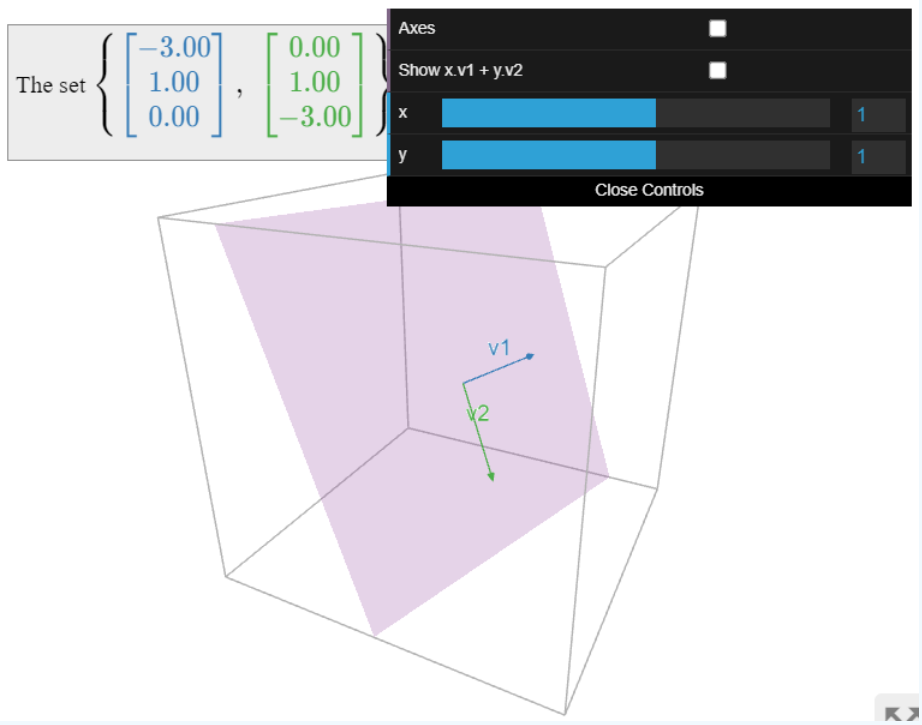


Figure 5.3.3: A picture of the plane V and its basis $\mathcal{B} = \{v_1, v_2\}$. Note that $\mathcal{B}$ spans V and is linearly independent.

This example is somewhat contrived, in that we will learn systematic methods for verifying that a subset is a basis. The intention is to illustrate the defining properties of a basis.

Computing a Basis for a Subspace

Now we show how to find bases for the column space of a matrix and the null space of a matrix. In order to find a basis for a given subspace, it is usually best to rewrite the subspace as a column space or a null space first: see this note in Section 2.6, Note 2.6.3

A Basis for the Column Space

First we show how to compute a basis for the column space of a matrix.

Theorem 5.3.1

The pivot columns of a matrix A form a basis for $\text{Col}(A)$.

Proof

This is a restatement of Theorem 2.5.3 in Section 2.5.

Note 5.3.2

The above theorem is referring to the pivot columns in the *original* matrix, not its reduced row echelon form. Indeed, a matrix and its reduced row echelon form generally have different column spaces. For example, in the matrix A below:

$$A = \begin{pmatrix} 1 & 2 & 0 & -1 \\ -2 & -3 & 4 & 5 \\ 2 & 4 & 0 & -2 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & -8 & -7 \\ 0 & 1 & 4 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

pivot columns = basis ⟷ pivot columns in RREF

Figure 5.3.4

the pivot columns are the first two columns, so a basis for $\mathrm{Col}(A)$ is

$$\left\{\begin{pmatrix}1\\-2\\2\end{pmatrix},\begin{pmatrix}2\\-3\\4\end{pmatrix}\right\}.$$

The first two columns of the reduced row echelon form certainly span a different subspace, as

$$\mathrm{Span}\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix}\right\}=\left\{\begin{pmatrix}a\\b\\0\end{pmatrix}\,|\,a,b\text{ in }\mathbb{R}\right\}=(x,y\text{-plane}),$$

but $\mathrm{Col}(A)$ contains vectors whose last coordinate is nonzero.

Corollary 5.3.1

The dimension of $\mathrm{Col}(A)$ is the number of pivots of A.

A Basis of a Span

Computing a basis for a span is the same as computing a basis for a column space. Indeed, the span of finitely many vectors $v_1,v_2,\ldots,v_m$ *is* the column space of a matrix, namely, the matrix A whose columns are $v_1,v_2,\ldots,v_m$:

$$A=\begin{pmatrix}|&|&&|\\v_1&v_2&\cdots&v_m\\|&|&&|\end{pmatrix}.$$

Example 5.3.6: A basis of a span

Find a basis of the subspace

$$V=\mathrm{Span}\left\{\begin{pmatrix}1\\-2\\2\end{pmatrix},\begin{pmatrix}2\\-3\\4\end{pmatrix},\begin{pmatrix}0\\4\\0\end{pmatrix},\begin{pmatrix}-1\\5\\-2\end{pmatrix}\right\}.$$

Solution

The subspace V is the column space of the matrix

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

The reduced row echelon form of this matrix is

$$\begin{pmatrix}1&0&-8&-7\\0&1&4&3\\0&0&0&0\end{pmatrix}.$$

The first two columns are pivot columns, so a basis for V is

$$\left\{\begin{pmatrix}1\\-2\\2\end{pmatrix},\begin{pmatrix}2\\-3\\4\end{pmatrix}\right\}.$$

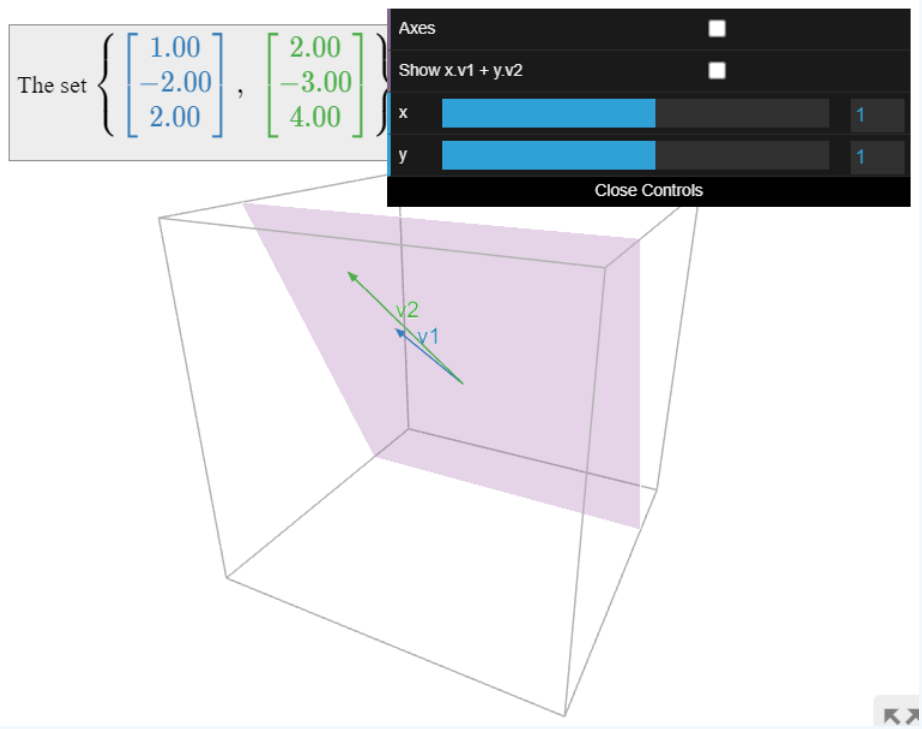


Figure 5.3.5: A picture of the plane V and its basis $\mathcal{B} = \{v_1, v_2\}$.

Example 5.3.7: Another basis of the same span

Find a basis of the subspace

$$V = \text{Span}\left\{\begin{pmatrix}1\\-2\\2\end{pmatrix}, \begin{pmatrix}2\\-3\\4\end{pmatrix}, \begin{pmatrix}0\\4\\0\end{pmatrix}, \begin{pmatrix}-1\\5\\-2\end{pmatrix}\right\}$$

which does not consist of the first two vectors, as in the previous Example 5.3.6.

Solution

The point of this example is that the above Theorem 5.3.1 gives *one* basis for V; as always, there are infinitely more.

Reordering the vectors, we can express V as the column space of

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

The reduced row echelon form of this matrix is

$$\begin{pmatrix}1 & 0 & 3/4 & 7/4\\0 & 1 & -1 & -2\\0 & 0 & 0 & 0\end{pmatrix}.$$

The first two columns are pivot columns, so a basis for V is

$$\left\{\begin{pmatrix}0\\4\\0\end{pmatrix}, \begin{pmatrix}-1\\5\\-2\end{pmatrix}\right\}.$$

These are the *last* two vectors in the given spanning set.

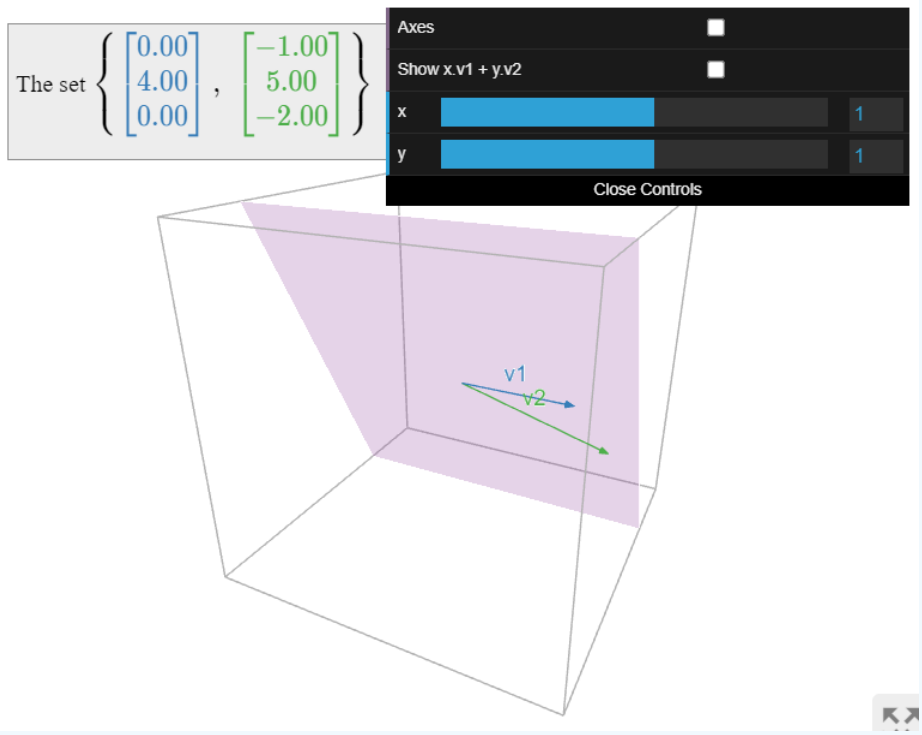


Figure 5.3.6: A picture of the plane V and its basis $\mathcal{B} = \{v_1, v_2\}$.

A Basis for the Null Space

In order to compute a basis for the null space of a matrix, one has to find the parametric vector form of the solutions of the homogeneous equation $Ax = 0$.

Theorem 5.3.2

The vectors attached to the free variables in the parametric vector form of the solution set of $Ax = 0$ form a basis of $\mathrm{Nul}(A)$.

The proof of the theorem has two parts. The first part is that every solution lies in the span of the given vectors. This is automatic: the vectors are exactly chosen so that every solution is a linear combination of those vectors. The second part is that the vectors are linearly independent. This part was discussed in Example 2.5.3 in Section 2.5.

A Basis for a General Subspace

As mentioned at the beginning of this subsection, when given a subspace written in a different form, in order to compute a basis it is usually best to rewrite it as a column space or null space of a matrix.

Example 5.3.8: A basis of a subspace

Let V be the subspace defined by

$$V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| x + 2y = z \right\}.$$

Find a basis for V. What is $\dim(V)$?

Solution

First we notice that V is exactly the solution set of the homogeneous linear equation $x + 2y - z = 0$. Hence $V = \mathrm{Nul}\begin{pmatrix} 1 & 2 & -1 \end{pmatrix}$. This matrix is in reduced row echelon form; the parametric form of the general solution is $x = -2y + z$, so the parametric vector form is

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = y \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

It follows that a basis is

$$\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Since V has a basis with two vectors, its dimension is 2: it is a plane.

The Basis Theorem

Recall that $\{v_1, v_2, \ldots, v_n\}$ forms a basis for $\mathbb{R}^n$ if and only if the matrix A with columns $v_1, v_2, \ldots, v_n$ has a pivot in every row and column (see this Example 5.3.4). Since A is an $n \times n$ matrix, these two conditions are equivalent: the vectors span if and only if they are linearly independent. The basis theorem is an abstract version of the preceding statement, that applies to any subspace.

Theorem 5.3.3: Basis Theorem

Let V be a subspace of dimension m. Then:

- Any m linearly independent vectors in V form a basis for V.
- Any m vectors that span V form a basis for V.

Proof

Suppose that $\mathcal{B} = \{v_1, v_2, \ldots, v_m\}$ is a set of linearly independent vectors in V. In order to show that $\mathcal{B}$ is a basis for V, we must prove that $V = \text{Span}\{v_1, v_2, \ldots, v_m\}$. If not, then there exists some vector v_{m+1} in V that is not contained in $\text{Span}\{v_1, v_2, \ldots, v_m\}$. By the increasing span criterion Theorem 2.5.2 in Section 2.5, the set $\{v_1, v_2, \ldots, v_m, v_{m+1}\}$ is also linearly independent. Continuing in this way, we keep choosing vectors until we eventually do have a linearly independent spanning set: say $V = \text{Span}\{v_1, v_2, \ldots, v_m, \ldots, v_{m+k}\}$. Then $\{v_1, v_2, \ldots, v_{m+k}\}$ is a basis for V, which implies that $\dim(V) = m + k > m$. But we were assuming that V has dimension m, so $\mathcal{B}$ must have already been a basis.

Now suppose that $\mathcal{B} = \{v_1, v_2, \ldots, v_m\}$ spans V. If $\mathcal{B}$ is not linearly independent, then by this Theorem 2.5.1 in Section 2.5, we can remove some number of vectors from $\mathcal{B}$ without shrinking its span. After reordering, we can assume that we removed the last k vectors without shrinking the span, and that we cannot remove any more. Now $V = \text{Span}\{v_1, v_2, \ldots, v_{m-k}\}$, and $\{v_1, v_2, \ldots, v_{m-k}\}$ is a basis for V because it is linearly independent. This implies that $\dim V = m - k < m$. But we were assuming that $\dim V = m$, so $\mathcal{B}$ must have already been a basis.

In other words, if you *already* know that $\dim V = m$, and if you have a set of m vectors $\mathcal{B} = \{v_1, v_2, \ldots, v_m\}$ in V, then you only have to check *one* of:

1. $\mathcal{B}$ is linearly independent, *or*
2. $\mathcal{B}$ spans V,

in order for $\mathcal{B}$ to be a basis of V. If you did not already know that $\dim V = m$, then you would have to check *both* properties.

To put it yet another way, suppose we have a set of vectors $\mathcal{B} = \{v_1, v_2, \ldots, v_m\}$ in a subspace V. Then if any two of the following statements is true, the third must also be true:

1. $\mathcal{B}$ is linearly independent,
2. $\mathcal{B}$ spans V, and
3. $\dim V = m$.

For example, if V is a plane, then any two noncollinear vectors in V form a basis.

Example 5.3.9: Two noncollinear vectors form a basis of a plane

Find a basis of the subspace

$$V = \text{Span}\left\{\begin{pmatrix}1\\-2\\2\end{pmatrix}, \begin{pmatrix}2\\-3\\4\end{pmatrix}, \begin{pmatrix}0\\4\\0\end{pmatrix}, \begin{pmatrix}-1\\5\\-2\end{pmatrix}\right\}$$

which is different from the bases in this Example 5.3.6 and this Example 5.3.7.

Solution

We know from the previous examples that $\dim V = 2$. By the Theorem 5.3.3, it suffices to find any two noncollinear vectors in V. We write two linear combinations of the four given spanning vectors, chosen at random:

$$w_1 = \begin{pmatrix}1\\-2\\2\end{pmatrix} + \begin{pmatrix}2\\-3\\4\end{pmatrix} = \begin{pmatrix}3\\-5\\6\end{pmatrix} \quad w_2 = -\begin{pmatrix}2\\-3\\4\end{pmatrix} + \frac{1}{2}\begin{pmatrix}0\\4\\0\end{pmatrix} = \begin{pmatrix}-2\\5\\-4\end{pmatrix}.$$

Since w_1, w_2 are not collinear, $\mathcal{B} = \{w_1, w_2\}$ is a basis for V.

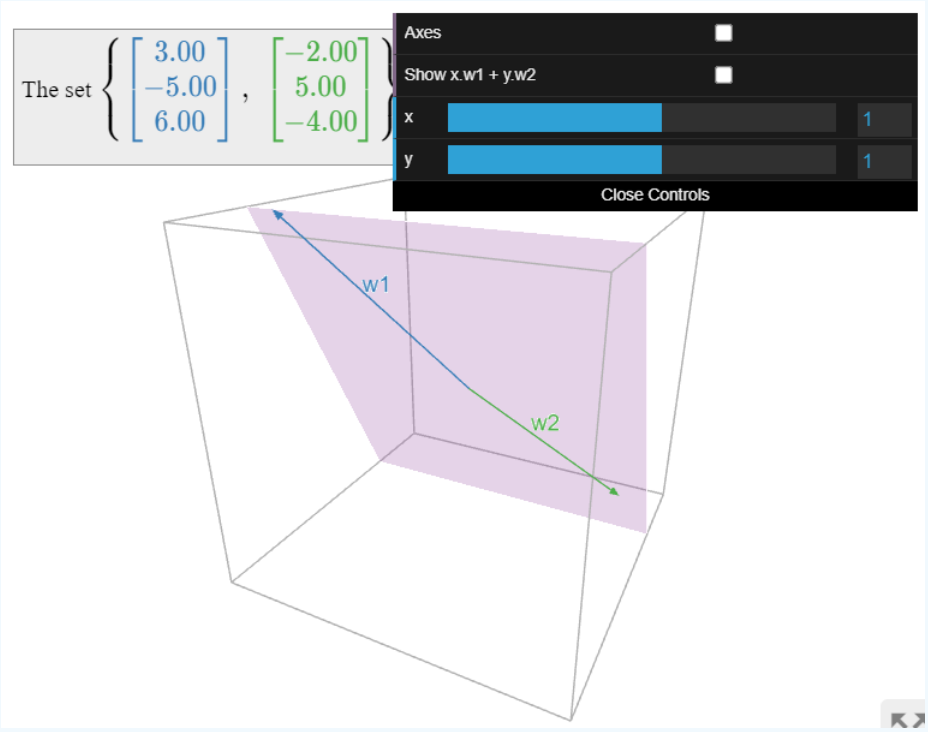


Figure 5.3.7: A picture of the plane V and its basis $\mathcal{B} = \{w_1, w_2\}$.

Example 5.3.10: Finding a basis by inspection

Find a basis for the plane

$$V = \left\{\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} \mid x_1 + x_2 = x_3\right\}$$

by inspection. (This plane is expressed in set builder notation, Note 2.2.3 in Section 2.2.)

Solution

First note that V is the null space of the matrix $\begin{pmatrix}1 & 1 & -1\end{pmatrix}$ this matrix is in reduced row echelon form and has two free variables, so V is indeed a plane. We write down two vectors satisfying $x_1 + x_2 = x_3$:

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

Since v_1 and v_2 are not collinear, they are linearly independent; since $\dim(V) = 2$, the basis theorem implies that $\{v_1, v_2\}$ is a basis for V.

This page titled 5.3: Basis and Dimension is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **2.7: Basis and Dimension** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

5.4: Rank and Homogeneous Systems

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 1.2.3. Our focus in this section is to consider what types of solutions are possible for a homogeneous system of equations.

Consider the following definition.

Consider the homogeneous system of equations given by

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

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 5.4.1: Solutions to a Homogeneous System of Equations

Find the nontrivial solutions to the following homogeneous system of equations

$$\begin{aligned} 2x + y - z &= 0 \\ x + 2y - 2z &= 0 \end{aligned}$$

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

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

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

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

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

$$x = 0$$
$$y = z = t$$
$$z = t$$

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

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

$$x = 0$$
$$y = 0 + t$$
$$z = 0 + t$$

can be written as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

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

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 5.4.1 has the basic solution $X_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

We explore this further in the following example.

Example 5.4.2: Basic Solutions of a Homogeneous System

Consider the following homogeneous system of equations.

$$x + 4y + 3z = 0$$
$$3x + 12y + 9z = 0$$

Find the basic solutions to this system.

Solution

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

$$\left[\begin{array}{ccc|c} 1 & 4 & 3 & 0 \\ 3 & 12 & 9 & 0 \end{array}\right] \to \cdots \to \left[\begin{array}{ccc|c} 1 & 4 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

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

$$x = -4s - 3t$$
$$y = s$$
$$z = t$$

which can be written as

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

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

$$X_1 = \begin{bmatrix} -4 \\ 1 \\ 0 \end{bmatrix}, X_2 = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$

We now present a new definition.

Definition 5.4.2: 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 5.4.2, this would also be a solution. For example, we could take the following linear combination

$$3 \begin{bmatrix} -4 \\ 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -18 \\ 3 \\ 2 \end{bmatrix}$$

You should take a moment to verify that

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -18 \\ 3 \\ 2 \end{bmatrix}$$

is in fact a solution to the system in Example 5.4.2.

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 5.4.3: Rank of a Matrix

The ***rank*** of a matrix A, written $\text{rank}(A)$, is the dimension of the column space $\text{Col}(A)$.

The ***nullity*** of a matrix A, written $\text{nullity}(A)$, is the dimension of the null space $\text{Nul}(A)$

Example 5.4.3: Finding the Rank of a Matrix

Consider the matrix

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

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

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

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 5.4.1: 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 5.4.2 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 5.4.1 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 5.4.2.

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 5.4.2: 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.

This page titled 5.4: Rank and Homogeneous Systems is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.5: Rank and Homogeneous Systems** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.
- **2.9: The Rank Theorem** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

5.5: The Rank Theorem

Objectives

1. Learn to understand and use the rank theorem.
2. *Picture:* the rank theorem.
3. *Theorem:* rank theorem.
4. *Vocabulary words:* ***rank***, ***nullity***.

In this section we present the rank theorem, which is the culmination of all of the work we have done so far.

The reader may have observed a relationship between the column space and the null space of a matrix. In Example 2.6.11 in Section 2.6, the column space and the null space of a 3×2 matrix are both lines, in $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively:

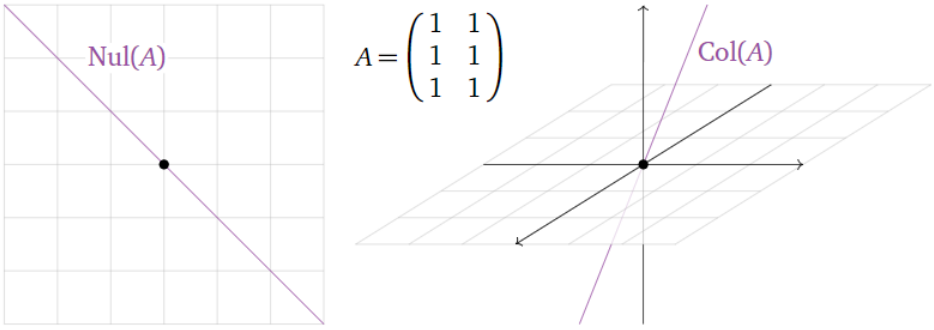


Figure 5.5.1

In Example 2.4.6 in Section 2.4, the null space of the 2×3 matrix

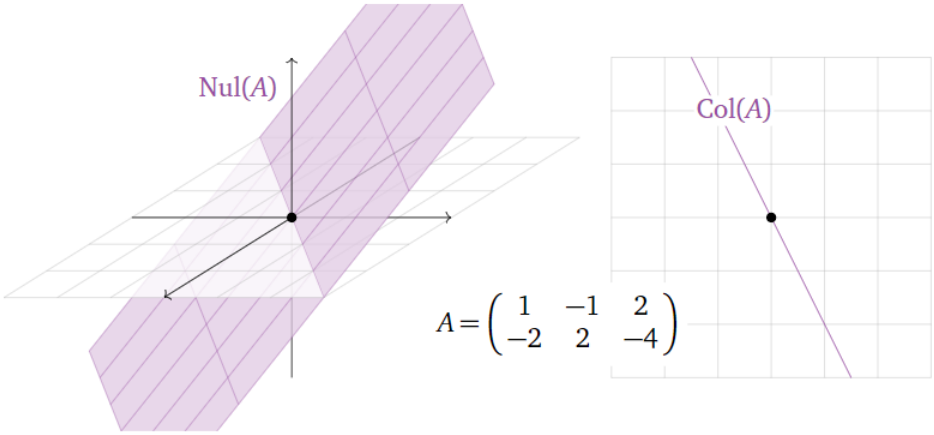


Figure 5.5.2

In Example 2.4.10 in Section 2.4, the null space of a 3×3 matrix is a line in $\mathbb{R}^3$, and the column space is a plane in $\mathbb{R}^3$:

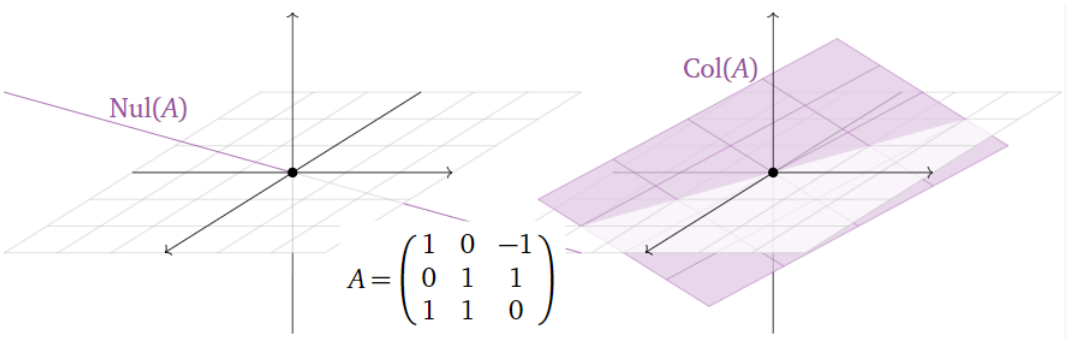


Figure 5.5.3

In all examples, the dimension of the column space plus the dimension of the null space is equal to the number of *columns* of the matrix. This is the content of the rank theorem.

Definition 5.5.1: Rank and Nullity

The ***rank*** of a matrix A, written $\text{rank}(A)$, is the dimension of the column space $\text{Col}(A)$.

The ***nullity*** of a matrix A, written $\text{nullity}(A)$, is the dimension of the null space $\text{Nul}(A)$.

The rank of a matrix A gives us important information about the solutions to $Ax = b$. Recall from Note 2.3.6 in Section 2.3 that $Ax = b$ is consistent exactly when b is in the span of the columns of A, in other words when b is in the column space of A. Thus, $\text{rank}(A)$ is the dimension of the set of b with the property that $Ax = b$ is consistent.

We know that the rank of A is equal to the number of pivot columns, Definition 1.2.5 in Section 1.2, (see this Theorem 2.7.1 in Section 2.7), and the nullity of A is equal to the number of free variables (see this Theorem 2.7.2 in Section 2.7), which is the number of columns without pivots. To summarize:

$$\begin{aligned} \text{rank}(A) = \dim \text{Col}(A) &= \text{the number of columns with pivots} \\ \text{nullity}(A) = \dim \text{Nul}(A) &= \text{the number of free variables} \\ &= \text{the number of columns without pivots} \end{aligned}$$

Clearly

$$\#(\text{columns with pivots}) + \#(\text{columns without pivots}) = \#(\text{columns}),$$

so we have proved the following theorem.

Theorem 5.5.1: Rank Theorem

If A is a matrix with n columns, then

$$\text{rank}(A) + \text{nullity}(A) = n.$$

In other words, for any consistent system of linear equations,

$$(\dim \text{ of column span}) + (\dim \text{ of solution set}) = (\text{number of variables}).$$

The rank theorem is really the culmination of this chapter, as it gives a strong relationship between the null space of a matrix (the solution set of $Ax = 0$) with the column space (the set of vectors b making $Ax = b$ consistent), our two primary objects of interest. The more freedom we have in choosing x the less freedom we have in choosing b and vice versa.

Example 5.5.1: Rank and nullity

Here is a concrete example of the rank theorem and the interplay between the degrees of freedom we have in choosing x and b in a matrix equation $Ax = b$.

Consider the matrices

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

If we multiply a vector (x, y, z) in $\mathbb{R}^3$ by A and B we obtain the vectors $Ax = (x, y, 0)$ and $Bx = (0, 0, z)$. So we can think of multiplication by A as giving the latitude and longitude of a point in $\mathbb{R}^3$ and we can think of multiplication by B as giving the height of a point in $\mathbb{R}^3$. The rank of A is 2 and the nullity is 1. Similarly, the rank of B is 1 and the nullity is 2.

These facts have natural interpretations. For the matrix A: the set of all latitudes and longitudes in $\mathbb{R}^3$ is a plane, and the set of points with the same latitude and longitude in $\mathbb{R}^3$ is a line; and for the matrix B: the set of all heights in $\mathbb{R}^3$ is a line, and the set of points at a given height in $\mathbb{R}^3$ is a plane. As the rank theorem tells us, we "trade off" having more choices for x for having more choices for b, and vice versa.

The rank theorem is a prime example of how we use the theory of linear algebra to say something qualitative about a system of equations without ever solving it. This is, in essence, the power of the subject.

Example 5.5.2: The rank is 2 and the nullity is 2

Consider the following matrix and its reduced row echelon form:

$$A = \begin{pmatrix} 1 & 2 & 0 & -1 \\ -2 & -3 & 4 & 5 \\ 2 & 4 & 0 & -2 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & -8 & -7 \\ 0 & 1 & 4 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

basis of Col(A) — free variables

Figure 5.5.4

A basis for $\text{Col}(A)$ is given by the pivot columns:

$$\left\{ \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix} \right\},$$

so $\text{rank}(A) = \dim \text{Col}(A) = 2$.

Since there are two free variables x_3, x_4, the null space of A has two vectors (see Theorem 2.7.2 in Section 2.7):

$$\left\{ \begin{pmatrix} 8 \\ -4 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 7 \\ -3 \\ 0 \\ 1 \end{pmatrix} \right\},$$

so $\text{nullity}(A) = 2$.

In this case, the rank theorem says that $2 + 2 = 4$, where 4 is the number of columns.

Example 5.5.3: Interactive: Rank is 1, nullity is 2

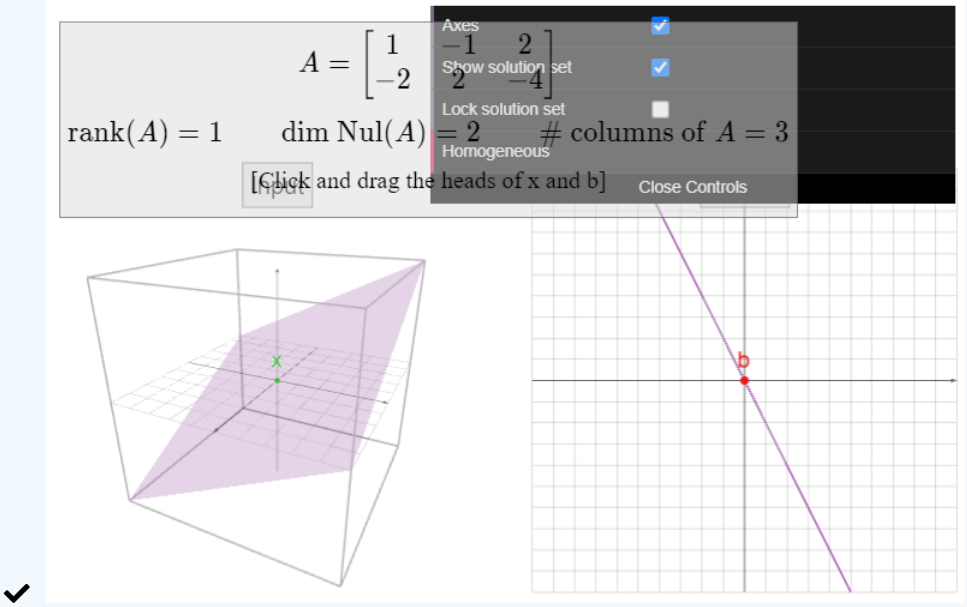


Figure 5.5.5: This 3×3 matrix has rank 1 and nullity 2. The violet plane on the left is the null space, and the violet line on the right is the column space.

Example 5.5.4: Interactive: Rank is 2, nullity is 1

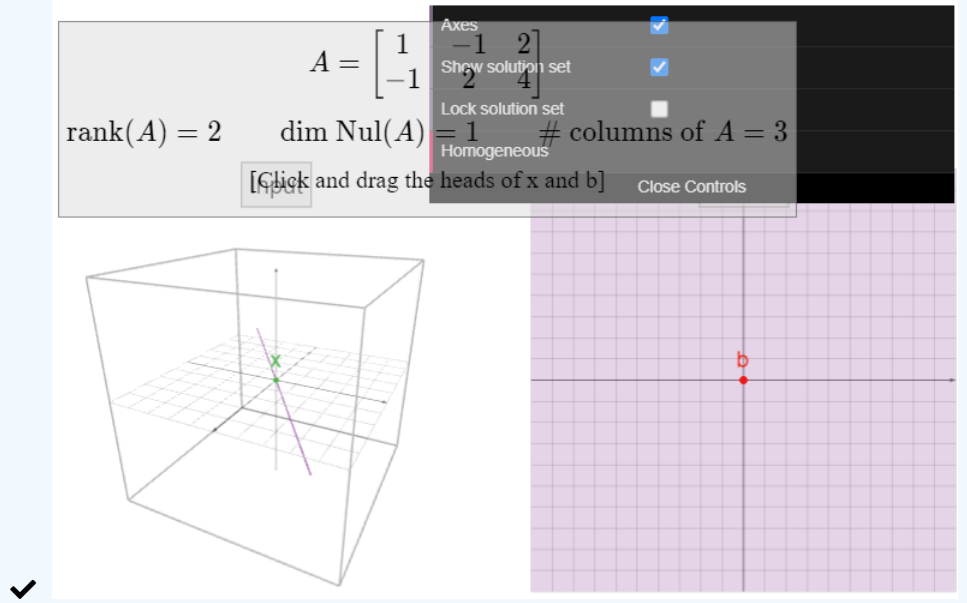


Figure 5.5.6: This 3×3 matrix has rank 2 and nullity 1. The violet line on the left is the null space, and the violet plane on the right is the column space.

This page titled 5.5: The Rank Theorem is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **2.9: The Rank Theorem** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

5.6: Bases as Coordinate Systems

Objectives

1. Learn to view a basis as a coordinate system on a subspace.
2. *Recipes:* compute the $\mathcal{B}$-coordinates of a vector, compute the usual coordinates of a vector from its $\mathcal{B}$-coordinates.
3. *Picture:* the $\mathcal{B}$-coordinates of a vector using its location on a nonstandard coordinate grid.
4. *Vocabulary word:* ***$\mathcal{B}$-coordinates***.

In this section, we interpret a basis of a subspace V as a *coordinate system* on V, and we learn how to write a vector in V in that coordinate system.

Fact 5.6.1

If $\mathcal{B} = \{v_1,\ v_2, \cdots, v_m\}$ is a basis for a subspace V, then any vector x in V can be written as a linear combination

$$x = c_1v_1 + c_2v_2 + \cdots + c_mv_m$$

in exactly one way.

Proof

Recall that to say $\mathcal{B}$ is a *basis* for V means that $\mathcal{B}$ spans V and $\mathcal{B}$ is linearly independent. Since $\mathcal{B}$ spans V, we can write any x in V as a linear combination of $v_1,\ v_2, \cdots, v_m$. For uniqueness, suppose that we had two such expressions:

$$\begin{aligned} x &= c_1v_1 + c_2v_2 + \cdots + c_mv_m \\ x &= c_1'v_1 + c_2'v_2 + \cdots + c_m'v_m. \end{aligned}$$

Subtracting the first equation from the second yields

$$0 = x - x = (c_1 - c_1')v_1 + (c_2 - c_2')v_2 + \cdots + (c_m - c_m')v_m.$$

Since $\mathcal{B}$ is linearly independent, the only solution to the above equation is the trivial solution: all the coefficients must be zero. It follows that $c_i - c_i'$ for all i, which proves that $c_1 = c_1',\ c_2 = c_2', \cdots, c_m = c_m'$.

Example 5.6.1

Consider the standard bases of $\mathbb{R}^3$ from Example 2.7.3 in Section 2.7.

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

According to the above fact, Fact 5.6.1, every vector in $\mathbb{R}^3$ can be written as a linear combination of e_1, e_2, e_3, with unique coefficients. For example,

$$v = \begin{pmatrix} 3 \\ 5 \\ -2 \end{pmatrix} = 3\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 5\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - 2\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 3e_1 + 5e_2 - 2e_3.$$

In this case, *the coordinates of v are exactly the coefficients of e_1, e_2, e_3.*

What exactly are coordinates, anyway? One way to think of coordinates is that they give directions for how to get to a certain point from the origin. In the above example, the linear combination $3e_1 + 5e_2 - 2e_3$ can be thought of as the following list of instructions: start at the origin, travel 3 units north, then travel 5 units east, then 2 units down.

Definition 5.6.1: $\mathcal{B}$-Coordinate Vector

Let $\mathcal{B} = \{v_1,\ v_2, \cdots, v_m\}$ be a basis of a subspace V, and let

$$x = c_1v_1 + c_2v_2 + \cdots + c_mv_m$$

be a vector in V. The coefficients $c_1,\ c_2, \cdots, c_m$ are the ***coordinates of x with respect to $\mathcal{B}$***. The ***$\mathcal{B}$-coordinate vector of x*** is the vector

$$[x]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} \quad \text{in } \mathbb{R}^m.$$

If we change the basis, then we can still give instructions for how to get to the point $(3,\ 5,\ -2)$, but the instructions will be different. Say for example we take the basis

$$v_1 = e_1 + e_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad v_2 = e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad v_3 = e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

We can write $(3,\ 5,\ -2)$ in this basis as $3v_1 + 2v_2 - 2v_3$. In other words: start at the origin, travel northeast 3 times as far as v_1, then 2 units east, then 2 units down. In this situation, we can say that "3 is the v_1-coordinate of $(3,\ 5,\ -2)$, 2 is the v_2-coordinate of $(3,\ 5,\ -2)$, and -2 is the v_3-coordinate of $(3,\ 5,\ -2)$."

Note 5.6.1

The above definition, Definition 5.6.1, gives a way of using $\mathbb{R}_m$ to *label* the points of a subspace of dimension m: a point is simply labeled by its $\mathcal{B}$-coordinate vector. For instance, if we choose a basis for a plane, we can label the points of that plane with the points of $\mathbb{R}^2$.

Example 5.6.2: A Nonstandard Coordinate System of $\mathbb{R}^2$)

Define

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \mathcal{B} = \{v_1,\ v_2\}.$$

1. Verify that $\mathcal{B}$ is a basis for $\mathbb{R}^2$.
2. If $[w]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, then what is w?
3. Find the $\mathcal{B}$-coordinates of $v = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$.

Solution

1. By the basis theorem in Section 2.7, Theorem 2.7.3, any two linearly independent vectors form a basis for $\mathbb{R}^2$. Clearly $v_1,\ v_2$ are not multiples of each other, so they are linearly independent.
2. To say $[w]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ means that 1 is the v_1-coordinate of w, and that 2 is the v_2-coordinate:

$$w = v_1 + 2v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 2\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}.$$

3. We have to solve the vector equation $v = c_1v_1 + c_2v_2$ in the unknowns $c_1,\ c_2$. We form an augmented matrix and row reduce:

$$\left(\begin{array}{cc|c} 1 & 1 & 5 \\ 1 & -1 & 3 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & 4 \\ 0 & 1 & 1 \end{array}\right).$$

We have $c_1 = 4$ and $c_2 = 1$, so $v = 4v_1 + v_2$ and $[v]_{\mathcal{B}} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$.

In the following picture, we indicate the coordinate system defined by $\mathcal{B}$ by drawing lines parallel to the "v_1-axis" and "v_2-axis". Using this grid it is easy to see that the $\mathcal{B}$-coordinates of v are $\begin{pmatrix} 5 \\ 1 \end{pmatrix}$, and that the $\mathcal{B}$-coordinates of w are $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

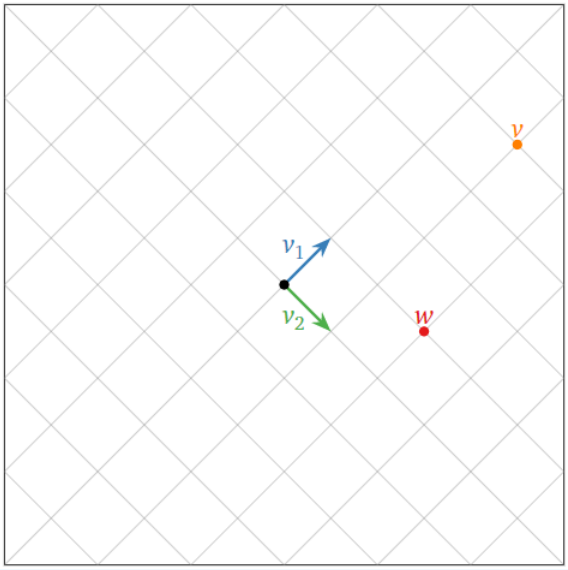


Figure 5.6.1

This picture could be the grid of streets in Palo Alto, California. Residents of Palo Alto refer to northwest as "north" and to northeast as "east". There is a reason for this: the old road to San Francisco is called El Camino Real, and that road runs from the southeast to the northwest in Palo Alto. So when a Palo Alto resident says "go south two blocks and east one block", they are giving directions from the origin to the Whole Foods at w.

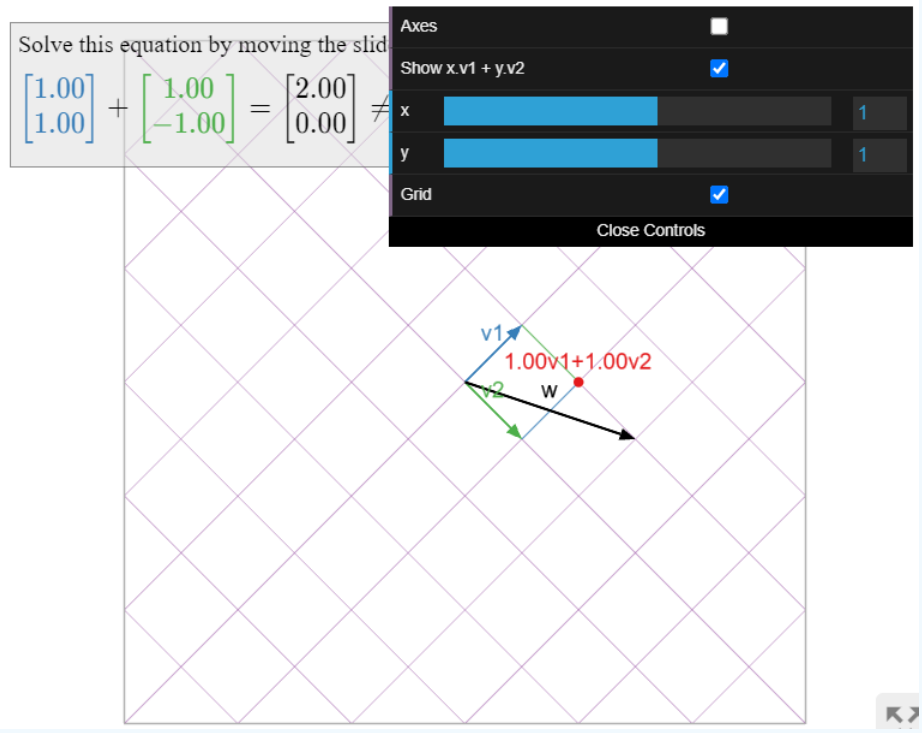


Figure 5.6.2: A picture of the basis $\mathcal{B} = \{v_1,\ v_2\}$ of $\mathbb{R}^2$. The grid indicates the coordinate system defined by the basis $\mathcal{B}$; one set of lines measures the v_1-coordinate, and the other set measures the v_2-coordinate. Use the sliders to find the $\mathcal{B}$-coordinates of w.

Example 5.6.3

Let

$$v_1 = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

These form a basis $\mathcal{B}$ for a plane $V = \text{Span}\{v_1,\ v_2\}$ in $\mathbb{R}^3$. We indicate the coordinate system defined by $\mathcal{B}$ by drawing lines parallel to the "v_1-axis" and "v_2-axis":

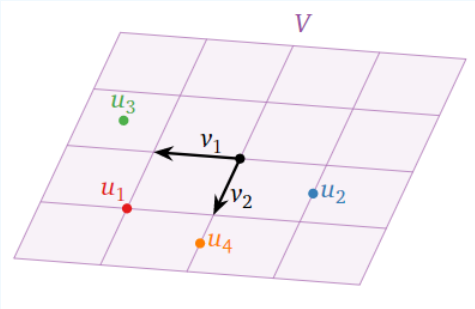


Figure 5.6.3

We can see from the picture that the v_1-coordinate of u_1 is equal to 1, as is the v_2-coordinate, so $[u_1]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Similarly, we have

$$[u_2]_{\mathcal{B}} = \begin{pmatrix} -1 \\ \frac{1}{2} \end{pmatrix} \quad [u_3]_{\mathcal{B}} = \begin{pmatrix} \frac{3}{2} \\ -\frac{1}{2} \end{pmatrix} \quad [u_4]_{\mathcal{B}} = \begin{pmatrix} 0 \\ \frac{3}{2} \end{pmatrix}.$$

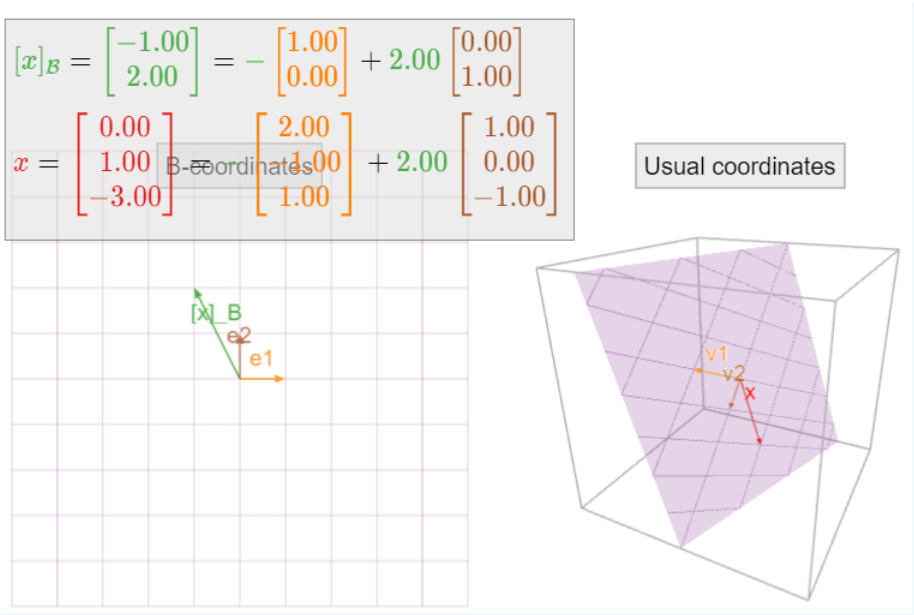


Figure 5.6.4: Left: the $\mathcal{B}$-coordinates of a vector x. Right: the vector x. The violet grid on the right is a picture of the coordinate system defined by the basis $\mathcal{B}$; one set of lines measures the v_1-coordinate, and the other set measures the v_2-coordinate. Drag the heads of the vectors x and $[x]_{\mathcal{B}}$ to understand the correspondence between x and its $\mathcal{B}$-coordinate vector.

Example 5.6.4: A Coordinate System on a Plane

Define

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},\ v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathcal{B} = \{v_1,\ v_2\}, \quad V = \text{Span}\{v_1,\ v_2\}.$$

1. Verify that $\mathcal{B}$ is a basis for V.
2. If $[w]_{\mathcal{B}} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$, then what is w?
3. Find the $\mathcal{B}$-coordinates of $v = \begin{pmatrix} 5 \\ 3 \\ 5 \end{pmatrix}$.

Solution

1. We need to verify that $\mathcal{B}$ spans V, and that it is linearly independent. By definition, V is the span of $\mathcal{B}$; since v_1 and v_2 are not multiples of each other, they are linearly independent. This shows in particular that V is a *plane*.
2. To say $[w]_{\mathcal{B}} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$ means that 5 is the v_1-coordinate of w, and that 2 is the v_2-coordinate:

$$w = 5v_1 + 2v_2 = 5\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + 2\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 2 \\ 7 \end{pmatrix}.$$

3. We have to solve the vector equation $v = c_1v_1 + c_2v_2$ in the unknowns $c_1,\ c_2$. We form an augmented matrix and row reduce:

$$\left(\begin{array}{cc|c} 1 & 1 & 5 \\ 0 & 1 & 3 \\ 1 & 1 & 5 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{array}\right).$$

We have $c_1 = 2$ and $c_2 = 3$, so $v = 2v_1 + 3v_2$ and $[v]_{\mathcal{B}} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

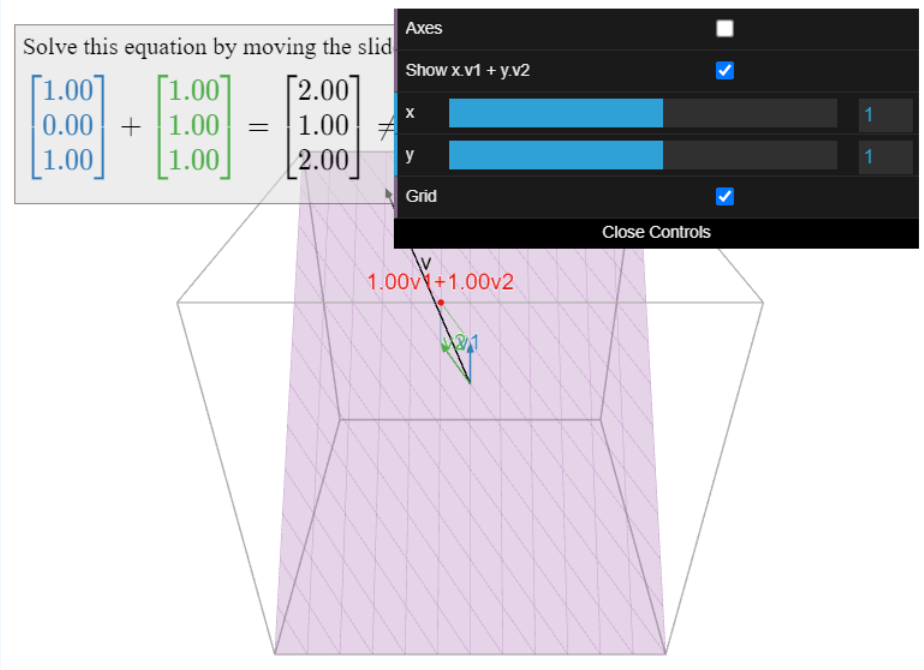


Figure 5.6.5: A picture of the plane V and the basis $\mathcal{B} = \{v_1,\ v_2\}$. The violet grid is a picture of the coordinate system defined by the basis $\mathcal{B}$; one set of lines measures the v_1-coordinate, and the other set measures the v_2-coordinate. Use the sliders to find the $\mathcal{B}$-coordinates of v.

Example 5.6.5: A Coordinate System on Another Plane

Define

$$v_1 = \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 2 \\ 8 \\ 6 \end{pmatrix}, \quad V = \text{Span}\{v_1,\ v_2,\ v_3\}.$$

1. Find a basis $\mathcal{B}$ for V.
2. Find the $\mathcal{B}$-coordinates of $x = \begin{pmatrix} 4 \\ 11 \\ 8 \end{pmatrix}$.

Solution

1. We write V as the column space of a matrix A, then row reduce to find the pivot columns, as in Example 2.7.6, in Section 2.7.

$$A = \begin{pmatrix} 2 & -1 & 2 \\ 3 & 1 & 8 \\ 2 & 1 & 6 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

The first two columns are pivot columns, so we can take $\mathcal{B} = \{v_1,\ v_2\}$ as our basis for V.

2. We have to solve the vector equation $x = c_1v_1 + c_2v_2$. We form an augmented matrix and row reduce:

$$\left(\begin{array}{cc|c} 2 & -1 & 4 \\ 3 & 1 & 11 \\ 2 & 1 & 8 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right).$$

We have $c_1 = 3$ and $c_2 = 2$, so $x = 3v_1 + 2v_2$, and thus $[x]_{\mathcal{B}} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$.

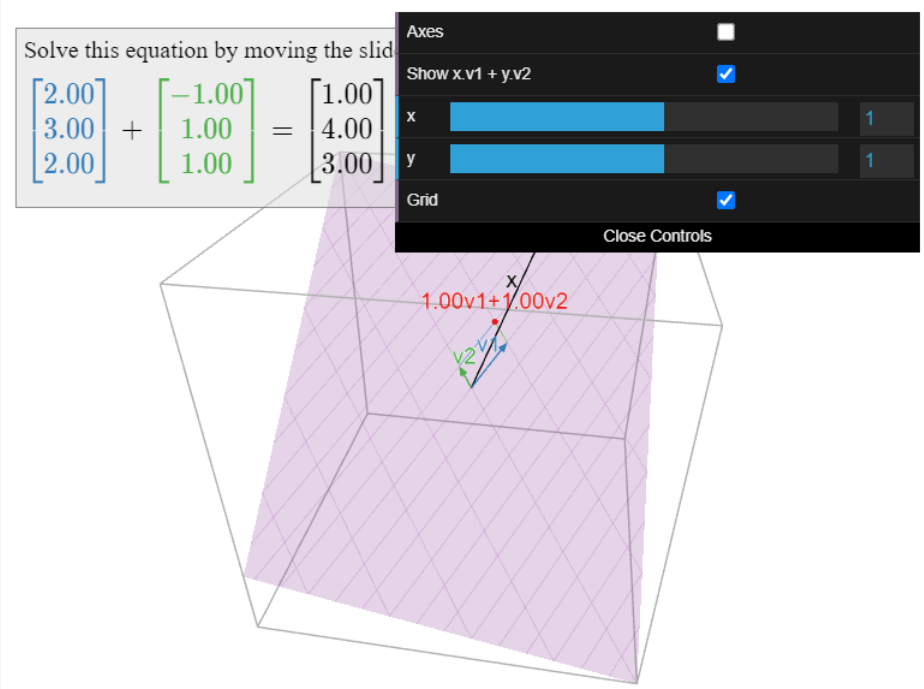


Figure 5.6.6: A picture of the plane V and the basis $\mathcal{B} = \{v_1,\ v_2\}$. The violet grid is a picture of the coordinate system defined by the basis $\mathcal{B}$; one set of lines measures the v_1-coordinate, and the other set measures the v_2-coordinate. Use the sliders to find the $\mathcal{B}$-coordinates of x.

Recipes: $\mathcal{B}$-Coordinates

If $\mathcal{B} = \{v_1,\ v_2, \cdots, v_m\}$ is a basis for a subspace V and x is in V, then

$$[x]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} \quad \text{means} \quad x = c_1v_1 + c_2v_2 + \cdots c_mv_m.$$

Finding the $\mathcal{B}$-coordinates of x means solving the vector equation

$$x = c_1v_1 + c_2v_2 + \cdots + c_mv_m$$

in the unknowns $c_1,\ c_2, \cdots, c_m$. This generally means row reducing the augmented matrix

$$\left(\begin{array}{cccc|c} | & | & & | & | \\ v_1 & v_2 & & v_m & x \\ | & | & & | & | \end{array}\right).$$

Remark

Let $\mathcal{B} = \{v_1,\ v_2, \cdots, v_m\}$ be a basis of a subspace V. Finding the $\mathcal{B}$-coordinates of a vector x means solving the vector equation

$$x = c_1v_1 + c_2v_2 + \cdots + c_mv_m.$$

> If x is *not* in V, then this equation has no solution, as x is not in $V = \text{Span}\{v_1, v_2, \cdots, v_m\}$. In other words, the above equation is *inconsistent* when x is not in V.

5.6: Bases as Coordinate Systems is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by LibreTexts.

- **2.8: Bases as Coordinate Systems** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

5.7: Spanning, Linear Independence and Basis in 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 5.7.1: 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

$$U \subseteq W$$

Spanning Set of Vectors

We begin this section with a definition.

Definition 5.7.2: 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 $\text{span}\{\vec{u}_1, \cdots, \vec{u}_k\}$.

Consider the following example.

Example 5.7.1: Span of Vectors

Describe the span of the vectors $\vec{u} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T$ and $\vec{v} = \begin{bmatrix} 3 & 2 & 0 \end{bmatrix}^T \in \mathbb{R}^3$.

Solution

You can see that any linear combination of the vectors $\vec{u}$ and $\vec{v}$ yields a vector of the form $\begin{bmatrix} x & y & 0 \end{bmatrix}^T$ 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 $\text{span}\{\vec{u}, \vec{v}\}$ 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 $\vec{w} = \begin{bmatrix} 4 & 5 & 0 \end{bmatrix}^T$. 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 $\vec{w} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$ 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 5.7.2: Vector in a Span

Let $\vec{u} = [\begin{array}{ccc} 1 & 1 & 0 \end{array}]^T$ and $\vec{v} = [\begin{array}{ccc} 3 & 2 & 0 \end{array}]^T \in \mathbb{R}^3$. Show that $\vec{w} = [\begin{array}{ccc} 4 & 5 & 0 \end{array}]^T$ is in $\text{span}\{\vec{u}, \vec{v}\}$.

Solution

For a vector to be in $\text{span}\{\vec{u}, \vec{v}\}$, it must be a linear combination of these vectors. If $\vec{w} \in \text{span}\{\vec{u}, \vec{v}\}$, we must be able to find scalars a, b such that

$$\vec{w} = a\vec{u} + b\vec{v}$$

We proceed as follows.

$$\begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix} = a \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}$$

This is equivalent to the following system of equations

$$\begin{aligned} a + 3b &= 4 \\ a + 2b &= 5 \end{aligned}$$

We solving this system the usual way, constructing the augmented matrix and row reducing to find the reduced row-echelon form.

$$\left[\begin{array}{cc|c} 1 & 3 & 4 \\ 1 & 2 & 5 \end{array}\right] \to \cdots \to \left[\begin{array}{cc|c} 1 & 0 & 7 \\ 0 & 1 & -1 \end{array}\right]$$

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 $\text{span}\{\vec{u}, \vec{v}\}$.

Linearly Independent Set of Vectors

We now turn our attention to the following question: what linear combinations of a given set of vectors $\{\vec{u}_1, \cdots, \vec{u}_k\}$ in $\mathbb{R}^n$ yields the zero vector? Clearly $0\vec{u}_1 + 0\vec{u}_2 + \cdots + 0\vec{u}_k = \vec{0}$, but is it possible to have $\sum_{i=1}^{k} a_i\vec{u}_i = \vec{0}$ without all coefficients being zero?

You can create examples where this easily happens. For example if $\vec{u}_1 = \vec{u}_2$, then $1\vec{u}_1 - \vec{u}_2 + 0\vec{u}_3 + \cdots + 0\vec{u}_k = \vec{0}$, no matter the vectors $\{\vec{u}_3, \cdots, \vec{u}_k\}$. 0But sometimes it can be more subtle.

Example 5.7.3: Linearly Dependent Set of Vectors

Consider the vectors

$$\vec{u}_1 = [\begin{array}{ccc} 0 & 1 & -2 \end{array}]^T, \vec{u}_2 = [\begin{array}{ccc} 1 & 1 & 0 \end{array}]^T, \vec{u}_3 = [\begin{array}{ccc} -2 & 3 & 2 \end{array}]^T, \text{ and } \vec{u}_4 = [\begin{array}{ccc} 1 & -2 & 0 \end{array}]^T$$

in $\mathbb{R}^3$.

Then verify that

$$1\vec{u}_1 + 0\vec{u}_2 + -\vec{u}_3 - 2\vec{u}_4 = \vec{0}$$

You can see that the linear combination does yield the zero vector but has some non-zero coefficients. Thus we define a set of vectors to be *linearly dependent* if this happens.

Definition 5.7.3: Linearly Dependent 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 dependent** if a linear combination of these vectors without all coefficients being zero does yield the zero vector.

Note that if $\sum_{i=1}^{k} a_i \vec{u}_i = \vec{0}$ and some coefficient is non-zero, say $a_1 \neq 0$, then

$$\vec{u}_1 = \frac{-1}{a_1} \sum_{i=2}^{k} a_i \vec{u}_i$$

and thus $\vec{u}_1$ is in the span of the other vectors. And the converse clearly works as well, so we get that a set of vectors is linearly dependent precisely when one of its vector is in the span of the other vectors of that set.

In particular, you can show that the vector $\vec{u}_1$ in the above example is in the span of the vectors $\{\vec{u}_2, \vec{u}_3, \vec{u}_4\}$.

If a set of vectors is NOT linearly dependent, then it must be that any linear combination of these vectors which yields the zero vector must use all zero coefficients. This is a very important notion, and we give it its own name of *linear independence*.

Definition 5.7.4: 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 whenever

$$\sum_{i=1}^{k} a_i \vec{u}_i = \vec{0}$$

it follows that each $a_i = 0$.

Note also that we require all vectors to be non-zero to form a linearly independent set.

To view this in a more familiar setting, form the $n \times k$ matrix A having these vectors as columns. Then all we are saying is that the set $\{\vec{u}_1, \cdots, \vec{u}_k\}$ is linearly independent precisely when $AX = 0$ has only the trivial solution.

Here is an example.

Example 5.7.4: Linearly Independent Vectors

Consider the vectors $\vec{u} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T$, $\vec{v} = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^T$, and $\vec{w} = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}^T$ in $\mathbb{R}^3$. Verify whether the set $\{\vec{u}, \vec{v}, \vec{w}\}$ is linearly independent.

Solution

So suppose that we have a linear combinations $a\vec{u} + b\vec{v} + c\vec{w} = \vec{0}$. Then you can see that this can only happen with $a = b = c = 0$.

As mentioned above, you can equivalently form the 3×3 matrix $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, and show that $AX = 0$ has only the trivial solution.

Thus this means the set $\{\vec{u}, \vec{v}, \vec{w}\}$ is linearly independent.

In terms of spanning, a set of vectors is linearly independent if it does not contain unnecessary vectors, that is not vector is in the span of the others.

Thus we put all this together in the following important theorem.

Theorem 5.7.1: Linear Independence as a Linear Combination

Let $\{\vec{u}_1, \cdots, \vec{u}_k\}$ be a collection of vectors in $\mathbb{R}^n$. Then the following are equivalent:

1. It is linearly independent, that is whenever

$$\sum_{i=1}^{k} a_i \vec{u}_i = \vec{0}$$

it follows that each coefficient $a_i = 0$.

2. No vector is in the span of the others.
3. 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.

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.

Here is a detailed example in $\mathbb{R}^4$.

Example 5.7.5: Linear Independence

Determine whether the set of vectors given by

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 2 \\ 0 \end{bmatrix} \right\}$$

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 5.7.6: Linear Independence

Determine whether the set of vectors given by

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 2 \\ -1 \end{bmatrix} \right\}$$

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:

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

Then by Theorem 5.7.1, 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

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

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

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

Therefore we can write

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

This can be rearranged as follows

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

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 5.7.7: 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 5.7.1: Linear Dependence in $\mathbb{R}''$

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 5.7.1.

Example 5.7.8: Linear Dependence

Consider the vectors

$$\left\{\begin{bmatrix}1\\4\end{bmatrix}, \begin{bmatrix}2\\3\end{bmatrix}, \begin{bmatrix}3\\2\end{bmatrix}\right\}$$

Are these vectors linearly independent?

Solution

This set contains three vectors in $\mathbb{R}^2$. By Corollary 5.7.1 these vectors are linearly dependent. In fact, we can write

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

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 5.7.2: Unique Linear Combination

Let $U\subseteq\mathbb{R}^n$ be an independent set. Then any vector $\vec{x}\in\text{span}(U)$ 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 $\vec{x}\in\text{span}(U)$ 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.Let $U\subseteq\mathbb{R}^n$ be an independent set. Then any vector $\vec{x}\in\text{span}(U)$ can be written uniquely as a linear combination of vectors of U.

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 $\vec{u}\notin\text{span}\{\vec{v}, \vec{w}\}$.

Example 5.7.9

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 $\vec{u} \notin \text{span}\{\vec{v}, \vec{w}\}$.

Solution

If $\vec{u} \in \text{span}\{\vec{v}, \vec{w}\}$, 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 $\vec{u} \notin \text{span}\{\vec{v}, \vec{w}\}$, 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 $\vec{u} \in \text{span}\{\vec{v}, \vec{w}\}$, 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 5.7.3: 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.

A Short Application to Chemistry

The following section applies the concepts of spanning and linear independence to the subject of chemistry.

When working with chemical reactions, there are sometimes a large number of reactions and some are in a sense redundant. Suppose you have the following chemical reactions.

$$CO + \tfrac{1}{2}O_2 \to CO_2$$
$$H_2 + \tfrac{1}{2}O_2 \to H_2O$$
$$CH_4 + \tfrac{3}{2}O_2 \to CO + 2H_2O$$
$$CH_4 + 2O_2 \to CO_2 + 2H_2O$$

There are four chemical reactions here but they are not independent reactions. There is some redundancy. What are the independent reactions? Is there a way to consider a shorter list of reactions? To analyze this situation, we can write the reactions in a matrix as follows

$$\begin{bmatrix} CO & O_2 & CO_2 & H_2 & H_2O & CH_4 \\ 1 & 1/2 & -1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1 & -1 & 0 \\ -1 & 3/2 & 0 & 0 & -2 & 1 \\ 0 & 2 & -1 & 0 & -2 & 1 \end{bmatrix}$$

Each row contains the coefficients of the respective elements in each reaction. For example, the top row of numbers comes from $CO + \frac{1}{2}O_2 - CO_2 = 0$ which represents the first of the chemical reactions.

We can write these coefficients in the following matrix

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

Rather than listing all of the reactions as above, it would be more efficient to only list those which are independent by throwing out that which is redundant. We can use the concepts of the previous section to accomplish this.

First, take the reduced row-echelon form of the above matrix.

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

The top three rows represent “independent" reactions which come from the original four reactions. One can obtain each of the original four rows of the matrix given above by taking a suitable linear combination of rows of this reduced row-echelon matrix.

With the redundant reaction removed, we can consider the simplified reactions as the following equations

$$CO + 3H_2 - 1H_2O - 1CH_4 = 0$$
$$O_2 + 2H_2 - 2H_2O = 0$$
$$CO_2 + 4H_2 - 2H_2O - 1CH_4 = 0$$

In terms of the original notation, these are the reactions

$$CO + 3H_2 \rightarrow H_2O + CH_4$$
$$O_2 + 2H_2 \rightarrow 2H_2O$$
$$CO_2 + 4H_2 \rightarrow 2H_2O + CH_4$$

These three reactions provide an equivalent system to the original four equations. The idea is that, in terms of what happens chemically, you obtain the same information with the shorter list of reactions. Such a simplification is especially useful when dealing with very large lists of reactions which may result from experimental evidence.

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 5.7.4: 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 5.7.10: 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 5.7.5: 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 5.7.5: 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 5.7.1, 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 $\text{span}\{\vec{u}_1, \cdots, \vec{u}_k\}$ and is therefore contained in V. This shows that $\text{span}\{\vec{u}_1, \cdots, \vec{u}_k\}$ 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 $a_1 \vec{u}_1 + a_2 \vec{u}_2 + \cdots + a_k \vec{u}_k \in W$.

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 5.7.2: 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 $V = \text{span}\{\vec{u}_1, \cdots, \vec{u}_k\}$.

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 5.7.6: 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. $\text{span}\{\vec{u}_1, \cdots, \vec{u}_k\} = V$
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 5.7.7: Standard Basis of $\mathbb{R}^n$

Let $\vec{e}_i$ be the vector in $\mathbb{R}^n$ which has a 1 in the ith entry and zeros elsewhere, that is the ith 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 5.7.6: 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 $\text{span}\{\vec{v}_1, \cdots, \vec{v}_s\}$ 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 $\text{span}\{\vec{v}_1, \cdots, \vec{v}_s\}$, 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 5.7.7: 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 9.4.4. 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 5.7.8: 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 5.7.3: 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 5.7.11: 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.

We continue by stating further properties of a set of vectors in $\mathbb{R}^n$.

Corollary 5.7.4: 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:

$$A = \begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_n \end{bmatrix}$$

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

$$\begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_m \end{bmatrix}$$

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 5.7.3

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

$$\begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_n \end{bmatrix}$$

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 5.7.3

The next theorem follows from the above claim.

Theorem 5.7.8: 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 5.7.4 together with Theorem 5.7.8. 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 5.7.12: Extending an Independent Set

Consider the set U given by

$$U = \left\{ \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \in \mathbb{R}^4 \middle| \ a - b = d - c \right\}$$

Then U is a subspace of $\mathbb{R}^4$ and $\dim(U) = 3$.

Then

$$S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 3 \\ 2 \end{bmatrix} \right\},$$

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 $\text{span}(S)$.

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

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

Therefore, S can be extended to the following basis of U:

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} \right\},$$

Next we consider the case of removing vectors from a spanning set to result in a basis.

Theorem 5.7.9: Finding a Basis from a Span

Let W be a subspace. Also suppose that $W = span\{\vec{w}_1, \cdots, \vec{w}_m\}$. 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 $\text{span}\{\vec{u}_1, \cdots, \vec{u}_k\} = W$. 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 5.7.13: Subset of a Span

Let W be the subspace

$$span\left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 8 \\ 19 \\ -8 \\ 8 \end{bmatrix}, \begin{bmatrix} -6 \\ -15 \\ 6 \\ -6 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \\ 0 \\ 1 \end{bmatrix} \right\}$$

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.

$$\begin{bmatrix} 1 & 1 & 8 & -6 & 1 & 1 \\ 2 & 3 & 19 & -15 & 3 & 5 \\ -1 & -1 & -8 & 6 & 0 & 0 \\ 1 & 1 & 8 & -6 & 1 & 1 \end{bmatrix}$$

Then take the reduced row-echelon form

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

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 , the same is true for the matrix having the given vectors as columns.

Consider the following theorems regarding a subspace contained in another subspace.

Theorem 5.7.10: 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 5.7.11: 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 5.7.14: Extending a Basis

Let $V = \mathbb{R}^4$ and let

$$W = \text{span}\left\{\begin{bmatrix}1\\0\\1\\1\end{bmatrix},\begin{bmatrix}0\\1\\0\\1\end{bmatrix}\right\}$$

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

$$\begin{bmatrix}1&0&1&0&0&0\\0&1&0&1&0&0\\1&0&0&0&1&0\\1&1&0&0&0&1\end{bmatrix} \tag{5.7.1}$$

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

$$\begin{bmatrix}1&0&0&0&1&0\\0&1&0&0&-1&1\\0&0&1&0&-1&0\\0&0&0&1&1&-1\end{bmatrix} \tag{5.7.2}$$

Therefore the pivot columns are

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

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 (5.7.1) obviously span $\mathbb{R}^4$. In fact the span of the first four is the same as the span of all six.

Consider another example.

Example 5.7.15: Extending a Basis

Let W be the span of $\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ in $\mathbb{R}^4$. Let V consist of the span of the vectors

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

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

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

Next find its reduced row-echelon form

$$\begin{bmatrix} 1 & 0 & 7 & -5 & 0 \\ 0 & 1 & -6 & 7 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

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.

Row Space, Column Space, and Null Space of a Matrix

We begin this section with a new definition.

Definition 5.7.9: Row and Column Space

Let A be an $m \times n$ matrix. The **column space** of A, written $\text{col}(A)$, is the span of the columns. The **row space** of A, written $\text{row}(A)$, 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 5.7.1: 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 $\text{row}(A) = \text{row}(B)$ [$\text{col}(A) = \text{col}(B)$].

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 $\text{row}(B) = \text{row}(A)$.
- 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

$$\text{row}(B) = \text{span}\{\vec{r}_1, \ldots, p\vec{r}_j, \ldots, \vec{r}_m\}.$$

Since

$$\{\vec{r}_1, \ldots, p\vec{r}_j, \ldots, \vec{r}_m\} \subseteq \text{row}(A),$$

it follows that $\text{row}(B) \subseteq \text{row}(A)$. Conversely, since

$$\{\vec{r}_1, \ldots, \vec{r}_m\} \subseteq \text{row}(B),$$

it follows that $\text{row}(A) \subseteq \text{row}(B)$. Therefore, $\text{row}(B) = \text{row}(A)$.

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

$$\text{row}(B) = \text{span}\{\vec{r}_1, \ldots, \vec{r}_{i-1}, \vec{r}_i + p\vec{r}_j, \ldots, \vec{r}_j, \ldots, \vec{r}_m\}.$$

Since

$$\{\vec{r}_1, \ldots, \vec{r}_{i-1}, \vec{r}_i + p\vec{r}_j, \ldots, \vec{r}_m\} \subseteq \text{row}(A),$$

it follows that $\text{row}(B) \subseteq \text{row}(A)$.

Conversely, since

$$\{\vec{r}_1, \ldots, \vec{r}_m\} \subseteq \text{row}(B),$$

it follows that $\text{row}(A) \subseteq \text{row}(B)$. Therefore, $\text{row}(B) = \text{row}(A)$.

Consider the following lemma.

Lemma 5.7.2: Row Space of a reduced row-echelon form Matrix

Let A be an $m \times n$ matrix and let R be its reduced row-echelon form. Then the nonzero rows of R form a basis of $\text{row}(R)$, and consequently of $\text{row}(A)$.

This lemma suggests that we can examine the reduced 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 reduced 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 $\text{col}(A)$.

Before proceeding to an example of this concept, we revisit the definition of rank.

Definition 5.7.10: Rank of a Matrix

Previously, we defined $\text{rank}(A)$ 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:

$$\text{rank}(A) = \dim(\text{row}(A))$$

Consider the following example.

Example 5.7.16: 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 5.7.2, we find the corresponding columns of A, in this case the first two columns. Therefore a basis for $\text{col}(A)$ 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 5.7.2 we know that the nonzero rows of R create a basis of $\text{row}(A)$. For the above matrix, the row space equals

$$\text{row}(A) = \text{span}\{[\,1 \quad 0 \quad -9 \quad 9 \quad 2\,], [\,0 \quad 1 \quad 5 \quad -3 \quad 0\,]\}$$

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 5.7.17: 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

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \frac{13}{2} \\ 0 & 1 & 0 & 2 & -\frac{5}{2} \\ 0 & 0 & 1 & -1 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and so the rank is 3. The row space is given by

$$\text{row}(A) = \text{span}\left\{ \begin{bmatrix} 1 & 0 & 0 & 0 & \frac{13}{2} \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 & 2 & -\frac{5}{2} \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & -1 & \frac{1}{2} \end{bmatrix} \right\}$$

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**,

$$\text{col}(A) = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \\ 1 \\ 2 \end{bmatrix} \right\}$$

Consider the solution given above for Example 5.7.17, 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 $\text{rank}(A) = \dim(\text{row}(A))$.

Theorem 5.7.12: Rank Theorem

Let A be an $m \times n$ matrix. Then $\dim(\text{col}(A))$, the dimension of the column space, is equal to the dimension of the row space, $\dim(\text{row}(A))$.

The following statements all follow from the Rank Theorem.

Corollary 5.7.5: Results of the Rank Theorem

Let A be a matrix. Then the following are true:

1. $\text{rank}(A) = \text{rank}(A^T)$.
2. For A of size $m \times n$, $\text{rank}(A) \leq m$ and $\text{rank}(A) \leq n$.
3. For A of size $n \times n$, A is invertible if and only if $\text{rank}(A) = n$.
4. For invertible matrices B and C of appropriate size, $\text{rank}(A) = \text{rank}(BA) = \text{rank}(AC)$.

Consider the following example.

Example 5.7.18: Rank of the Transpose

Let

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

Find $\text{rank}(A)$ and $\text{rank}(A^T)$.

Solution

To find $\text{rank}(A)$ we first row reduce to find the reduced row-echelon form.

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

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.

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

You can see that $\text{rank}(A^T) = 2$, the same as $\text{rank}(A)$.

We now define what is meant by the null space of a general $m \times n$ matrix.

Definition 5.7.11: 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.

$$\text{null}\,(A) = \left\{ \vec{x} : A\vec{x} = \vec{0} \right\}$$

It can also be referred to using the notation $\ker(A)$. Similarly, we can discuss the image of A, denoted by $\text{im}\,(A)$. The image of A consists of the vectors of $\mathbb{R}^m$ which "get hit" by A. The formal definition is as follows.

Definition 5.7.12: Image of A

The image of A, written $\text{im}\,(A)$ is given by

$$\text{im}\,(A) = \{ A\vec{x} : \vec{x} \in \mathbb{R}^n \}$$

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.

$$\begin{matrix} \text{null}(A) & & \text{im}(A) \\ \mathbb{R}^n & \overset{A}{\to} & \mathbb{R}^m \end{matrix}$$

As indicated, $\text{im}\,(A)$ is a subset of $\mathbb{R}^m$ while $\text{null}\,(A)$ 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 5.7.19: Null Space

Let A be an $m \times n$ matrix. Then the null space of A, $\text{null}(A)$ is a subspace of $\mathbb{R}^n$.

Solution

- Since $A\vec{0}_n = \vec{0}_m$, $\vec{0}_n \in \text{null}(A)$.
- Let $\vec{x}, \vec{y} \in \text{null}(A)$. 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 $\vec{x} + \vec{y} \in \text{null}(A)$.
- Let $\vec{x} \in \text{null}(A)$ 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 $k\vec{x} \in \text{null}(A)$.

Therefore by the subspace test, $\text{null}(A)$ is a subspace of $\mathbb{R}^n$.

The proof that $\text{im}(A)$ 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 $\text{null}(A)$ for a matrix A. However, finding $\text{null}\,(A)$ is not new! There is just some new terminology being used, as $\text{null}\,(A)$ is simply the solution to the system $A\vec{x} = \vec{0}$.

Theorem 5.7.13: Basis of null(A)

Let A be an $m \times n$ matrix such that $\text{rank}(A) = r$. Then the system $A\vec{x} = \vec{0}_m$ has $n - r$ basic solutions, providing a basis of $\text{null}(A)$ with $\dim(\text{null}(A)) = n - r$.

Consider the following example.

Example 5.7.20: Null Space of A

Let

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

Find $\text{null}\,(A)$ and $\text{im}\,(A)$.

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

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

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

$$\begin{bmatrix} -3t \\ t \\ t \end{bmatrix} : t \in \mathbb{R}$$

which can be written as

$$t \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} : t \in \mathbb{R}$$

Therefore, the null space of A is all multiples of this vector, which we can write as

$$\text{null}(A) = \text{span}\left\{ \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \right\}$$

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

$$\text{im}\,(A) = \text{col}(A) = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \right\}$$

Here is a larger example, but the method is entirely similar.

Example 5.7.21: Null Space of A

Let

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

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

$$\left[\begin{array}{ccccc|c} 1 & 2 & 1 & 0 & 1 & 0 \\ 2 & -1 & 1 & 3 & 0 & 0 \\ 3 & 1 & 2 & 3 & 1 & 0 \\ 4 & -2 & 2 & 6 & 0 & 0 \end{array}\right] \to \cdots \to \left[\begin{array}{ccccc|c} 1 & 0 & \frac{3}{5} & \frac{6}{5} & \frac{1}{5} & 0 \\ 0 & 1 & \frac{1}{5} & -\frac{3}{5} & \frac{2}{5} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

It follows that the first two columns are pivot columns, and the next three correspond to parameters. Therefore, $\text{null}\,(A)$ 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

$$s\begin{bmatrix} -\frac{3}{5} \\ -\frac{1}{5} \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} -\frac{6}{5} \\ \frac{3}{5} \\ 0 \\ 1 \\ 0 \end{bmatrix} + r\begin{bmatrix} \frac{1}{5} \\ -\frac{2}{5} \\ 0 \\ 0 \\ 1 \end{bmatrix} : s,t,r \in \mathbb{R}.$$

In other words, the null space of this matrix equals the span of the three vectors above. Thus

$$\text{null}\,(A) = \text{span}\left\{\begin{bmatrix} -\frac{3}{5} \\ -\frac{1}{5} \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{6}{5} \\ \frac{3}{5} \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{5} \\ -\frac{2}{5} \\ 0 \\ 0 \\ 1 \end{bmatrix}\right\}$$

Notice also that the three vectors above are linearly independent and so the dimension of $\text{null}\,(A)$ 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 form equals 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 5.7.13: Nullity

The dimension of the null space of a matrix is called the **nullity**, denoted $\dim(\text{null}\,(A))$.

From our observation above we can now state an important theorem.

Theorem 5.7.14: Rank and Nullity

Let A be an $m \times n$ matrix. Then $\text{rank}\,(A) + \dim(\text{null}\,(A)) = n$.

Example 5.7.22: Rank and Nullity

Let

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

Find $\text{rank}\,(A)$ and $\dim(\text{null}\,(A))$.

Solution

In the above Example 5.7.20 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 5.7.14 that $\text{rank}\,(A) + \dim(\text{null}\,(A)) = 2 + 1 = 3$, which is the number of columns of A.

We conclude this section with two similar, and important, theorems.

Theorem 5.7.15

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

1. $\text{rank}(A) = n$.
2. $\text{row}(A) = \mathbb{R}^n$, i.e., the rows of A span $\mathbb{R}^n$.
3. The columns of A are independent in $\mathbb{R}^m$.
4. The $n \times n$ matrix $A^T A$ is invertible.
5. There exists an $n \times m$ matrix C so that $CA = I_n$.
6. If $A\vec{x} = \vec{0}_m$ for some $\vec{x} \in \mathbb{R}^n$, then $\vec{x} = \vec{0}_n$.

Theorem 5.7.16

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 $m \times m$ matrix AA^T is invertible.
5. There exists an $n \times m$ matrix C so that $AC = I_m$.

6. The system $A\vec{x} = \vec{b}$ is consistent for every $\vec{b} \in \mathbb{R}^m$.

This page titled 5.7: Spanning, Linear Independence and Basis in R^n is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **4.10: Spanning, Linear Independence and Basis in R^n** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

5.8: A Linear Transformations

Outcomes

A. Understand the definition of a linear transformation in the context of vector spaces.

Recall that a function is simply a transformation of a vector to result in a new vector. Consider the following definition.

Definition 5.8.1: Linear Transformation

Let V and W be vector spaces. Suppose $T : V \mapsto W$ is a function, where for each $\vec{x} \in V, T(\vec{x}) \in W$. Then T is a **linear transformation** if whenever k, p are scalars and $\vec{v}_1$ and $\vec{v}_2$ are vectors in V

$$T(k\vec{v}_1 + p\vec{v}_2) = kT(\vec{v}_1) + pT(\vec{v}_2)$$

Several important examples of linear transformations include the zero transformation, the identity transformation, and the scalar transformation.

Example 5.8.1: Linear Transformations

Let V and W be vector spaces.

1. **The zero transformation**
   $0 : V \to W$ is defined by $0(\vec{v}) = \vec{0}$ for all $\vec{v} \in V$.
2. **The identity transformation**
   $1_V : V \to V$ is defined by $1_V(\vec{v}) = \vec{v}$ for all $\vec{v} \in V$.
3. **The scalar transformation** Let $a \in \mathbb{R}$.
   $s_a : V \to V$ is defined by $s_a(\vec{v}) = a\vec{v}$ for all $\vec{v} \in V$.

Solution

We will show that the scalar transformation s_a is linear, the rest are left as an exercise.

By Definition 5.8.1 we must show that for all scalars k, p and vectors $\vec{v}_1$ and $\vec{v}_2$ in V, $s_a(k\vec{v}_1 + p\vec{v}_2) = ks_a(\vec{v}_1) + ps_a(\vec{v}_2)$. Assume that a is also a scalar.

$$\begin{aligned} s_a(k\vec{v}_1 + p\vec{v}_2) &= a(k\vec{v}_1 + p\vec{v}_2) \\ &= ak\vec{v}_1 + ap\vec{v}_2 \\ &= k(a\vec{v}_1) + p(a\vec{v}_2) \\ &= ks_a(\vec{v}_1) + ps_a(\vec{v}_2) \end{aligned}$$

Therefore s_a is a linear transformation.

Consider the following important theorem.

Theorem 5.8.1: Properties of Linear Transformations

Let V and W be vector spaces, and $T : V \mapsto W$ a linear transformation. Then

1. T preserves the zero vector.
$$T(\vec{0}) = \vec{0}$$
2. T preserves additive inverses. For all $\vec{v} \in V$,
$$T(-\vec{v}) = -T(\vec{v})$$
3. T preserves linear combinations. For all $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m \in V$ and all $k_1, k_2, \ldots, k_m \in \mathbb{R}$,
$$T(k_1\vec{v}_1 + k_2\vec{v}_2 + \cdots + k_m\vec{v}_m) = k_1T(\vec{v}_1) + k_2T(\vec{v}_2) + \cdots + k_mT(\vec{v}_m).$$

Proof

1. Let $\vec{0}_V$ denote the zero vector of V and let $\vec{0}_W$ denote the zero vector of W. We want to prove that $T(\vec{0}_V) = \vec{0}_W$. Let $\vec{v} \in V$. Then $0\vec{v} = \vec{0}_V$ and
$$T(\vec{0}_V) = T(0\vec{v}) = 0T(\vec{v}) = \vec{0}_W.$$
2. Let $\vec{v} \in V$; then $-\vec{v} \in V$ is the additive inverse of $\vec{v}$, so $\vec{v} + (-\vec{v}) = \vec{0}_V$. Thus
$$\begin{aligned} T(\vec{v} + (-\vec{v})) &= T(\vec{0}_V) \\ T(\vec{v}) + T(-\vec{v})) &= \vec{0}_W \\ T(-\vec{v}) &= \vec{0}_W - T(\vec{v}) = -T(\vec{v}). \end{aligned}$$
3. This result follows from preservation of addition and preservation of scalar multiplication. A formal proof would be by induction on m.

Consider the following example using the above theorem.

Example 5.8.2: Linear Combination

Let $T : \mathbb{P}_2 \to \mathbb{R}$ be a linear transformation such that
$$T(x^2 + x) = -1; T(x^2 - x) = 1; T(x^2 + 1) = 3.$$
Find $T(4x^2 + 5x - 3)$.

We provide two solutions to this problem.

Solution 1:

Suppose $a(x^2 + x) + b(x^2 - x) + c(x^2 + 1) = 4x^2 + 5x - 3$. Then
$$(a + b + c)x^2 + (a - b)x + c = 4x^2 + 5x - 3.$$
Solving for a, b, and c results in the unique solution $a = 6$, $b = 1$, $c = -3$. Thus
$$\begin{aligned} T(4x^2 + 5x - 3) &= T(6(x^2 + x) + (x^2 - x) - 3(x^2 + 1)) \\ &= 6T(x^2 + x) + T(x^2 - x) - 3T(x^2 + 1) \\ &= 6(-1) + 1 - 3(3) = -14. \end{aligned}$$

Solution 2:

Notice that $S = \{x^2 + x, x^2 - x, x^2 + 1\}$ is a basis of $\mathbb{P}_2$, and thus x^2, x, and 1 can each be written as a linear combination of elements of S.
$$\begin{aligned} x^2 &= \tfrac{1}{2}(x^2 + x) + \tfrac{1}{2}(x^2 - x) \\ x &= \tfrac{1}{2}(x^2 + x) - \tfrac{1}{2}(x^2 - x) \\ 1 &= (x^2 + 1) - \tfrac{1}{2}(x^2 + x) - \tfrac{1}{2}(x^2 - x). \end{aligned}$$
Then
$$\begin{aligned} T(x^2) &= T\left(\tfrac{1}{2}(x^2 + x) + \tfrac{1}{2}(x^2 - x)\right) = \tfrac{1}{2}T(x^2 + x) + \tfrac{1}{2}T(x^2 - x) \\ &= \tfrac{1}{2}(-1) + \tfrac{1}{2}(1) = 0. \\ T(x) &= T\left(\tfrac{1}{2}(x^2 + x) - \tfrac{1}{2}(x^2 - x)\right) = \tfrac{1}{2}T(x^2 + x) - \tfrac{1}{2}T(x^2 - x) \\ &= \tfrac{1}{2}(-1) - \tfrac{1}{2}(1) = -1. \\ T(1) &= T\left((x^2 + 1) - \tfrac{1}{2}(x^2 + x) - \tfrac{1}{2}(x^2 - x)\right) \\ &= T(x^2 + 1) - \tfrac{1}{2}T(x^2 + x) - \tfrac{1}{2}T(x^2 - x) \\ &= 3 - \tfrac{1}{2}(-1) - \tfrac{1}{2}(1) = 3. \end{aligned}$$

Therefore,

$$\begin{aligned} T(4x^2+5x-3) &= 4T(x^2)+5T(x)-3T(1) \\ &= 4(0)+5(-1)-3(3)=-14. \end{aligned}$$

The advantage of **Solution 2** over **Solution 1** is that if you were now asked to find $T(-6x^2-13x+9)$, it is easy to use $T(x^2)=0$, $T(x)=-1$ and $T(1)=3$:

$$\begin{aligned} T(-6x^2-13x+9) &= -6T(x^2)-13T(x)+9T(1) \\ &= -6(0)-13(-1)+9(3)=13+27=40. \end{aligned}$$

More generally,

$$\begin{aligned} T(ax^2+bx+c) &= aT(x^2)+bT(x)+cT(1) \\ &= a(0)+b(-1)+c(3)=-b+3c. \end{aligned}$$

Suppose two linear transformations act in the same way on $\vec{v}$ for all vectors. Then we say that these transformations are equal.

Definition 5.8.2: Equal Transformations

Let S and T be linear transformations from V to W. Then $S=T$ if and only if for every $\vec{v}\in V$,

$$S(\vec{v})=T(\vec{v})$$

The definition above requires that two transformations have the same action on every vector in order for them to be equal. The next theorem argues that it is only necessary to check the action of the transformations on basis vectors.

Theorem 5.8.2: Transformation of a Spanning Set

Let V and W be vector spaces and suppose that S and T are linear transformations from V to W. Then in order for S and T to be equal, it suffices that $S(\vec{v}_i)=T(\vec{v}_i)$ where $V=span\{\vec{v}_1,\vec{v}_2,\ldots,\vec{v}_n\}$.

This theorem tells us that a linear transformation is completely determined by its actions on a spanning set. We can also examine the effect of a linear transformation on a basis.

Theorem 5.8.3: Transformation of a Basis

Suppose V and W are vector spaces and let $\{\vec{w}_1,\vec{w}_2,\ldots,\vec{w}_n\}$ be any given vectors in W that may not be distinct. Then there exists a basis $\{\vec{v}_1,\vec{v}_2,\ldots,\vec{v}_n\}$ of V and a unique linear transformation $T:V\mapsto W$ with $T(\vec{v}_i)=\vec{w}_i$.

Furthermore, if

$$\vec{v}=k_1\vec{v}_1+k_2\vec{v}_2+\cdots+k_n\vec{v}_n$$

is a vector of V, then

$$T(\vec{v})=k_1\vec{w}_1+k_2\vec{w}_2+\cdots+k_n\vec{w}_n.$$

This page titled 5.8: A Linear Transformations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **9.6: Linear Transformations** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

5.9: A Isomorphisms

Outcomes

A. Apply the concepts of one to one and onto to transformations of vector spaces.
B. Determine if a linear transformation of vector spaces is an isomorphism.
C. Determine if two vector spaces are isomorphic.

One to One and Onto Transformations

Recall the following definitions, given here in terms of vector spaces.

Definition 5.9.1: One to One Transformation

Let V, W be vector spaces with $\vec{v}_1, \vec{v}_2$ vectors in V. Then a linear transformation $T : V \mapsto W$ is called **one to one** if whenever $\vec{v}_1 \neq \vec{v}_2$ it follows that

$$T(\vec{v}_1) \neq T(\vec{v}_2)$$

Definition 5.9.2: Onto Transformation

Let V, W be vector spaces. Then a linear transformation $T : V \mapsto W$ is called **onto** if for all $\vec{w} \in \vec{W}$ there exists $\vec{v} \in V$ such that $T(\vec{v}) = \vec{w}$.

Recall that every linear transformation T has the property that $T(\vec{0}) = \vec{0}$. This will be necessary to prove the following useful lemma.

Lemma 5.9.1: One to One

The assertion that a linear transformation T is one to one is equivalent to saying that if $T(\vec{v}) = \vec{0}$, then $\vec{v} = 0$.

Proof

Suppose first that T is one to one.

$$T(\vec{0}) = T\left(\vec{0} + \vec{0}\right) = T(\vec{0}) + T(\vec{0})$$

and so, adding the additive inverse of $T(\vec{0})$ to both sides, one sees that $T(\vec{0}) = \vec{0}$. Therefore, if $T(\vec{v}) = \vec{0}$, it must be the case that $\vec{v} = \vec{0}$ because it was just shown that $T(\vec{0}) = \vec{0}$.

Now suppose that if $T(\vec{v}) = \vec{0}$, then $\vec{v} = 0$. If $T(\vec{v}) = T(\vec{u})$, then $T(\vec{v}) - T(\vec{u}) = T(\vec{v} - \vec{u}) = \vec{0}$ which shows that $\vec{v} - \vec{u} = 0$ or in other words, $\vec{v} = \vec{u}$.

Consider the following example.

Example 5.9.1: One to One Transformation

Let $S : \mathbb{P}_2 \to \mathbb{M}_{22}$ be a linear transformation defined by

$$S(ax^2 + bx + c) = \begin{bmatrix} a+b & a+c \\ b-c & b+c \end{bmatrix}$$

for all $ax^2 + bx + c \in \mathbb{P}_2$.

Prove that S is one to one but not onto.

Solution

By definition,

$$\ker(S) = \{ax^2 + bx + c \in \mathbb{P}_2 \mid a+b=0, a+c=0, b-c=0, b+c=0\}.$$

Suppose $p(x) = ax^2 + bx + c \in \ker(S)$. This leads to a homogeneous system of four equations in three variables. Putting the augmented matrix in reduced row-echelon form:

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

The solution is $a = b = c = 0$. This tells us that if $S(p(x)) = 0$, then $p(x) = ax^2 + bx + c = 0x^2 + 0x + 0 = 0$. Therefore it is one to one.

To show that S is **not** onto, find a matrix $A \in \mathbb{M}_{22}$ such that for every $p(x) \in \mathbb{P}_2$, $S(p(x)) \neq A$. Let

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

and suppose $p(x) = ax^2 + bx + c \in \mathbb{P}_2$ is such that $S(p(x)) = A$. Then

$$\begin{array}{ll} a+b=0 & a+c=1 \\ b-c=0 & b+c=2 \end{array}$$

Solving this system

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

Since the system is inconsistent, there is no $p(x) \in \mathbb{P}_2$ so that $S(p(x)) = A$, and therefore S is not onto.

Example 5.9.2: An Onto Transformation

Let $T : \mathbb{M}_{22} \to \mathbb{R}^2$ be a linear transformation defined by

$$T\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+d \\ b+c \end{bmatrix} \text{ for all } \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{M}_{22}.$$

Prove that T is onto but not one to one.

Solution

Let $\begin{bmatrix} x \\ y \end{bmatrix}$ be an arbitrary vector in $\mathbb{R}^2$. Since $T\begin{bmatrix} x & y \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$, T is onto.

By Lemma 5.9.1 T is one to one if and only if $T(A) = \vec{0}$ implies that $A = 0$ the zero matrix. Observe that

$$T\left(\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\right) = \begin{bmatrix} 1+-1 \\ 0+0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

There exists a nonzero matrix A such that $T(A) = \vec{0}$. It follows that T is not one to one.

The following example demonstrates that a one to one transformation preserves linear independence.

Example 5.9.3: One to One and Independence

Let V and W be vector spaces and $T : V \mapsto W$ a linear transformation. Prove that if T is one to one and $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is an independent subset of V, then $\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_k)\}$ is an independent subset of W.

Solution

Let $\vec{0}_V$ and $\vec{0}_W$ denote the zero vectors of V and W, respectively. Suppose that

$$a_1 T(\vec{v}_1) + a_2 T(\vec{v}_2) + \cdots + a_k T(\vec{v}_k) = \vec{0}_W$$

for some $a_1, a_2, \ldots, a_k \in \mathbb{R}$. Since linear transformations preserve linear combinations (addition and scalar multiplication),

$$T(a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_k\vec{v}_k) = \vec{0}_W.$$

Now, since T is one to one, $\ker(T) = \{\vec{0}_V\}$, and thus

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_k\vec{v}_k = \vec{0}_V.$$

However, $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is independent so $a_1 = a_2 = \cdots = a_k = 0$. Therefore, $\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_k)\}$ is independent.

A similar claim can be made regarding onto transformations. In this case, an onto transformation preserves a spanning set.

Example 5.9.4: Onto and Spanning

Let V and W be vector spaces and $T : V \to W$ a linear transformation. Prove that if T is onto and $V = span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$, then

$$W = span\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_k)\}.$$

Solution

Suppose that T is onto and let $\vec{w} \in W$. Then there exists $\vec{v} \in V$ such that $T(\vec{v}) = \vec{w}$. Since $V = span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$, there exist $a_1, a_2, \ldots a_k \in \mathbb{R}$ such that $\vec{v} = a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_k\vec{v}_k$. Using the fact that T is a linear transformation,

$$\begin{aligned} \vec{w} = T(\vec{v}) &= T(a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_k\vec{v}_k) \\ &= a_1 T(\vec{v}_1) + a_2 T(\vec{v}_2) + \cdots + a_k T(\vec{v}_k), \end{aligned}$$

i.e., $\vec{w} \in span\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_k)\}$, and thus

$$W \subseteq span\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_k)\}.$$

Since $T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_k) \in W$, it follows from that $span\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_k)\} \subseteq W$, and therefore $W = span\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_k)\}$.

Isomorphisms

The focus of this section is on linear transformations which are both one to one and onto. When this is the case, we call the transformation an isomorphism.

Definition 5.9.3: Isomorphism

Let V and W be two vector spaces and let $T : V \mapsto W$ be a linear transformation. Then T is called an **isomorphism** if the following two conditions are satisfied.

- T is one to one.
- T is onto.

Definition 5.9.4: Isomorphic

Let V and W be two vector spaces and let $T : V \mapsto W$ be a linear transformation. Then if T is an isomorphism, we say that V and W are **isomorphic.**

Consider the following example of an isomorphism.

Example 5.9.5: Isomorphism

Let $T : \mathbb{M}_{22} \to \mathbb{R}^4$ be defined by

$$T\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \text{ for all } \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{M}_{22}.$$

Show that T is an isomorphism.

Solution

Notice that if we can prove T is an isomorphism, it will mean that $\mathbb{M}_{22}$ and $\mathbb{R}^4$ are isomorphic. It remains to prove that

1. T is a linear transformation;
2. T is one-to-one;
3. T is onto.

***T* is linear:** Let k, p be scalars.

$$\begin{aligned} T\left(k\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} + p\begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix}\right) &= T\left(\begin{bmatrix} ka_1 & kb_1 \\ kc_1 & kd_1 \end{bmatrix} + \begin{bmatrix} pa_2 & pb_2 \\ pc_2 & pd_2 \end{bmatrix}\right) \\ &= T\left(\begin{bmatrix} ka_1 + pa_2 & kb_1 + pb_2 \\ kc_1 + pc_2 & kd_1 + pd_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} ka_1 + pa_2 \\ kb_1 + pb_2 \\ kc_1 + pc_2 \\ kd_1 + pd_2 \end{bmatrix} \\ &= \begin{bmatrix} ka_1 \\ kb_1 \\ kc_1 \\ kd_1 \end{bmatrix} + \begin{bmatrix} pa_2 \\ pb_2 \\ pc_2 \\ pd_2 \end{bmatrix} \\ &= k\begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ d_1 \end{bmatrix} + p\begin{bmatrix} a_2 \\ b_2 \\ c_2 \\ d_2 \end{bmatrix} \\ &= kT\left(\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix}\right) + pT\left(\begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix}\right) \end{aligned}$$

Therefore T is linear.

***T* is one-to-one:** By Lemma 5.9.1 we need to show that if $T(A) = 0$ then $A = 0$ for some matrix $A \in \mathbb{M}_{22}$.

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

This clearly only occurs when $a = b = c = d = 0$ which means that

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

Hence T is one-to-one.

***T* is onto:** Let

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \in \mathbb{R}^4,$$

and define matrix $A \in \mathbb{M}_{22}$ as follows:

$$A = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}.$$

Then $T(A) = \vec{x}$, and therefore T is onto.

Since T is a linear transformation which is one-to-one and onto, T is an isomorphism. Hence $\mathbb{M}_{22}$ and $\mathbb{R}^4$ are isomorphic.

An important property of isomorphisms is that the inverse of an isomorphism is itself an isomorphism and the composition of isomorphisms is an isomorphism. We first recall the definition of composition.

Definition 5.9.5: Composition of Transformations

Let V, W, Z be vector spaces and suppose $T : V \mapsto W$ and $S : W \mapsto Z$ are linear transformations. Then the composite of S and T is

$$S \circ T : V \mapsto Z$$

and is defined by

$$(S \circ T)(\vec{v}) = S(T(\vec{v})) \text{ for all } \vec{v} \in V$$

Consider now the following proposition.

Proposition 5.9.1: Composite and Inverse Isomorphism

Let $T : V \to W$ be an isomorphism. Then $T^{-1} : W \to V$ is also an isomorphism. Also if $T : V \to W$ is an isomorphism and if $S : W \to Z$ is an isomorphism for the vector spaces V, W, Z, then $S \circ T$ defined by $(S \circ T)(v) = S(T(v))$ is also an isomorphism.

Proof

Consider the first claim. Since T is onto, a typical vector in W is of the form $T(\vec{v})$ where $\vec{v} \in V$. Consider then for a, b scalars,

$$T^{-1}\left(aT(\vec{v}_1) + bT(\vec{v}_2)\right)$$

where $\vec{v}_1, \vec{v}_2 \in V$. Consider if this is equal to

$$aT^{-1}\left(T(\vec{v}_1)\right) + bT^{-1}\left(T(\vec{v}_2)\right) = a\vec{v}_1 + b\vec{v}_2?$$

Since T is one to one, this will be so if

$$T\left(a\vec{v}_1 + b\vec{v}_2\right) = T\left(T^{-1}\left(aT(\vec{v}_1) + bT(\vec{v}_2)\right)\right) = aT(\vec{v}_1) + bT(\vec{v}_2)$$

However, the above statement is just the condition that T is a linear map. Thus T^{-1} is indeed a linear map. If $\vec{v} \in V$ is given, then $\vec{v} = T^{-1}\left(T(\vec{v})\right)$ and so T^{-1} is onto. If $T^{-1}(\vec{v}) = \vec{0}$, then

$$\vec{v} = T\left(T^{-1}(\vec{v})\right) = T(\vec{0}) = \vec{0}$$

and so T^{-1} is one to one.

Next suppose T and S are as described. Why is $S \circ T$ a linear map? Let for a, b scalars,

$$\begin{aligned} S \circ T\left(a\vec{v}_1 + b\vec{v}_2\right) &\equiv S\left(T\left(a\vec{v}_1 + b\vec{v}_2\right)\right) = S\left(aT(\vec{v}_1) + bT(\vec{v}_2)\right) \\ &= aS\left(T(\vec{v}_1)\right) + bS\left(T(\vec{v}_2)\right) \equiv a\left(S \circ T\right)(\vec{v}_1) + b\left(S \circ T\right)(\vec{v}_2) \end{aligned}$$

Hence $S \circ T$ is a linear map. If $\left(S \circ T\right)(\vec{v}) = 0$, then $S\left(T\left(\vec{v}\right)\right) = \vec{0}$ and it follows that $T(\vec{v}) = \vec{0}$ and hence by this lemma again, $\vec{v} = \vec{0}$. Thus $S \circ T$ is one to one. It remains to verify that it is onto. Let $\vec{z} \in Z$. Then since S is onto, there exists $\vec{w} \in W$ such that $S(\vec{w}) = \vec{z}$. Also, since T is onto, there exists $\vec{v} \in V$ such that $T(\vec{v}) = \vec{w}$. It follows that $S\left(T\left(\vec{v}\right)\right) = \vec{z}$ and so $S \circ T$ is also onto.

Suppose we say that two vector spaces V and W are related if there exists an isomorphism of one to the other, written as $V \sim W$. Then the above proposition suggests that $\sim$ is an equivalence relation. That is: $\sim$ satisfies the following conditions:

- $V \sim V$
- If $V \sim W$, it follows that $W \sim V$
- If $V \sim W$ and $W \sim Z$, then $V \sim Z$

We leave the proof of these to the reader.

The following fundamental lemma describes the relation between bases and isomorphisms.

Lemma 5.9.2: Bases and Isomorphisms

Let $T : V \to W$ be a linear map where V, W are vector spaces. Then a linear transformation T which is one to one has the property that if $\{\vec{u}_1, \cdots, \vec{u}_k\}$ is linearly independent, then so is $\{T(\vec{u}_1), \cdots, T(\vec{u}_k)\}$. More generally, T is an isomorphism if and only if whenever $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is a basis for V, it follows that $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ is a basis for W.

Proof

First suppose that T is a linear map and is one to one and $\{\vec{u}_1, \cdots, \vec{u}_k\}$ is linearly independent. It is required to show that $\{T(\vec{u}_1), \cdots, T(\vec{u}_k)\}$ is also linearly independent. Suppose then that

$$\sum_{i=1}^{k} c_i T(\vec{u}_i) = \vec{0}$$

Then, since T is linear,

$$T\left(\sum_{i=1}^{n} c_i \vec{u}_i\right) = \vec{0}$$

Since T is one to one, it follows that

$$\sum_{i=1}^{n} c_i \vec{u}_i = 0$$

Now the fact that $\{\vec{u}_1, \cdots, \vec{u}_n\}$ is linearly independent implies that each $c_i = 0$. Hence $\{T(\vec{u}_1), \cdots, T(\vec{u}_n)\}$ is linearly independent.

Now suppose that T is an isomorphism and $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is a basis for V. It was just shown that $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ is linearly independent. It remains to verify that the span of $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ is all of W. This is where T is onto is used. If $\vec{w} \in W$, there exists $\vec{v} \in V$ such that $T(\vec{v}) = \vec{w}$. Since $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is a basis, it follows that there exists scalars $\{c_i\}_{i=1}^{n}$ such that

$$\sum_{i=1}^{n} c_i \vec{v}_i = \vec{v}.$$

Hence,

$$\vec{w} = T(\vec{v}) = T\left(\sum_{i=1}^{n} c_i \vec{v}_i\right) = \sum_{i=1}^{n} c_i T\vec{v}_i$$

which shows that the span of these vectors $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ is all of W showing that this set of vectors is a basis for W.

Next suppose that T is a linear map which takes a basis to a basis. Then for $\{\vec{v}_1, \cdots, \vec{v}_n\}$ a basis for V, it follows $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ is a basis for W. Then if $w \in W$, there exist scalars c_i such that $w = \sum_{i=1}^{n} c_i T(\vec{v}_i) = T\left(\sum_{i=1}^{n} c_i \vec{v}_i\right)$ showing that T is onto. If $T\left(\sum_{i=1}^{n} c_i \vec{v}_i\right) = 0$ then $\sum_{i=1}^{n} c_i T(\vec{v}_i) = \vec{0}$ and since the vectors $\{T(\vec{v}_1), \cdots, T(\vec{v}_n)\}$ are linearly independent, it follows that each $c_i = 0$. Since $\sum_{i=1}^{n} c_i \vec{v}_i$ is a typical vector in V, this has shown that if $T(\vec{v}) = 0$ then $\vec{v} = \vec{0}$ and so T is also one to one. Thus T is an isomorphism.

The following theorem illustrates a very useful idea for defining an isomorphism. Basically, if you know what it does to a basis, then you can construct the isomorphism.

Theorem 5.9.1: Isomorphic Vector Spaces

Suppose V and W are two vector spaces. Then the two vector spaces are isomorphic if and only if they have the same dimension. In the case that the two vector spaces have the same dimension, then for a linear transformation $T : V \to W$, the following are equivalent.

1. T is one to one.
2. T is onto.
3. T is an isomorphism.

Proof

Suppose first these two vector spaces have the same dimension. Let a basis for V be $\{\vec{v}_1, \cdots, \vec{v}_n\}$ and let a basis for W be $\{\vec{w}_1, \cdots, \vec{w}_n\}$. Now define T as follows.

$$T(\vec{v}_i) = \vec{w}_i$$

for $\sum_{i=1}^{n} c_i \vec{v}_i$ an arbitrary vector of V,

$$T\left(\sum_{i=1}^{n} c_i \vec{v}_i\right) = \sum_{i=1}^{n} c_i T(\vec{v}_i) = \sum_{i=1}^{n} c_i \vec{w}_i.$$

It is necessary to verify that this is well defined. Suppose then that

$$\sum_{i=1}^{n} c_i \vec{v}_i = \sum_{i=1}^{n} \hat{c}_i \vec{v}_i$$

Then

$$\sum_{i=1}^{n} (c_i - \hat{c}_i)\, \vec{v}_i = 0$$

and since $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is a basis, $c_i = \hat{c}_i$ for each i. Hence

$$\sum_{i=1}^{n} c_i \vec{w}_i = \sum_{i=1}^{n} \hat{c}_i \vec{w}_i$$

and so the mapping is well defined. Also if a, b are scalars,

$$\begin{aligned} T\left(a\sum_{i=1}^{n}c_i\vec{v}_i+b\sum_{i=1}^{n}\hat{c}_i\vec{v}_i\right) &= T\left(\sum_{i=1}^{n}\left(ac_i+b\hat{c}_i\right)\vec{v}_i\right)=\sum_{i=1}^{n}\left(ac_i+b\hat{c}_i\right)\vec{w}_i \\ &= a\sum_{i=1}^{n}c_i\vec{w}_i+b\sum_{i=1}^{n}\hat{c}_i\vec{w}_i \\ &= aT\left(\sum_{i=1}^{n}c_i\vec{v}_i\right)+bT\left(\sum_{i=1}^{n}\hat{c}_i\vec{v}_i\right) \end{aligned}$$

Thus T is a linear map.

Now if

$$T\left(\sum_{i=1}^{n}c_i\vec{v}_i\right)=\sum_{i=1}^{n}c_i\vec{w}_i=\vec{0},$$

then since the $\{\vec{w}_1,\cdots,\vec{w}_n\}$ are independent, each $c_i=0$ and so $\sum_{i=1}^{n}c_i\vec{v}_i=\vec{0}$ also. Hence T is one to one. If $\sum_{i=1}^{n}c_i\vec{w}_i$ is a vector in W, then it equals

$$\sum_{i=1}^{n}c_iT\vec{v}_i=T\left(\sum_{i=1}^{n}c_i\vec{v}_i\right)$$

showing that T is also onto. Hence T is an isomorphism and so V and W are isomorphic.

Next suppose these two vector spaces are isomorphic. Let T be the name of the isomorphism. Then for $\{\vec{v}_1,\cdots,\vec{v}_n\}$ a basis for V, it follows that a basis for W is $\{T\vec{v}_1,\cdots,T\vec{v}_n\}$ showing that the two vector spaces have the same dimension.

Now suppose the two vector spaces have the same dimension.

First consider the claim that $1.)\Rightarrow 2.)$. If T is one to one, then if $\{\vec{v}_1,\cdots,\vec{v}_n\}$ is a basis for V, then $\{T(\vec{v}_1),\cdots,T(\vec{v}_n)\}$ is linearly independent. If it is not a basis, then it must fail to span W. But then there would exist $\vec{w}\notin span\{T(\vec{v}_1),\cdots,T(\vec{v}_n)\}$ and it follows that $\{T(\vec{v}_1),\cdots,T(\vec{v}_n),\vec{w}\}$ would be linearly independent which is impossible because there exists a basis for W of n vectors. Hence

$$span\{T(\vec{v}_1),\cdots,T(\vec{v}_n)\}=W$$

and so $\{T(\vec{v}_1),\cdots,T(\vec{v}_n)\}$ is a basis. Hence, if $\vec{w}\in W$, there exist scalars c_i such that

$$\vec{w}=\sum_{i=1}^{n}c_iT(\vec{v}_i)=T\left(\sum_{i=1}^{n}c_i\vec{v}_i\right)$$

showing that T is onto. This shows that $1.)\Rightarrow 2.)$.

Next consider the claim that $2.)\Rightarrow 3.)$. Since $2.)$ holds, it follows that T is onto. It remains to verify that T is one to one. Since T is onto, there exists a basis of the form $\{T(\vec{v}_i),\cdots,T(\vec{v}_n)\}$. If $\{\vec{v}_1,\cdots,\vec{v}_n\}$ is linearly independent, then this set of vectors must also be a basis for V because if not, there would exist $\vec{u}\notin span\{\vec{v}_1,\cdots,\vec{v}_n\}$ so $\{\vec{v}_1,\cdots,\vec{v}_n,\vec{u}\}$ would be a linearly independent set which is impossible because by assumption, there exists a basis which has n vectors. So why is $\{\vec{v}_1,\cdots,\vec{v}_n\}$ linearly independent? Suppose

$$\sum_{i=1}^{n}c_i\vec{v}_i=\vec{0}$$

Then

$$\sum_{i=1}^{n}c_iT\vec{v}_i=\vec{0}$$

Hence each $c_i=0$ and so, as just discussed, $\{\vec{v}_1,\cdots,\vec{v}_n\}$ is a basis for V. Now it follows that a typical vector in V is of the form $\sum_{i=1}^{n}c_i\vec{v}_i$. If $T\left(\sum_{i=1}^{n}c_i\vec{v}_i\right)=\vec{0}$, it follows that

$$\sum_{i=1}^{n} c_i T(\vec{v}_i) = \vec{0}$$

and so, since $\{T(\vec{v}_i), \cdots, T(\vec{v}_n)\}$ is independent, it follows each $c_i = 0$ and hence $\sum_{i=1}^{n} c_i \vec{v}_i = \vec{0}$. Thus T is one to one as well as onto and so it is an isomorphism.

If T is an isomorphism, it is both one to one and onto by definition so $3.)$ implies both $1.)$ and $2.)$.

Note the interesting way of defining a linear transformation in the first part of the argument by describing what it does to a basis and then "extending it linearly".

Consider the following example.

Example 5.9.6

Let $V = \mathbb{R}^3$ and let W denote the polynomials of degree at most 2. Show that these two vector spaces are isomorphic.

Solution

First, observe that a basis for W is $\{1, x, x^2\}$ and a basis for V is $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$. Since these two have the same dimension, the two are isomorphic. An example of an isomorphism is this:

$$T(\vec{e}_1) = 1, T(\vec{e}_2) = x, T(\vec{e}_3) = x^2$$

and extend T linearly as in the above proof. Thus

$$T(a, b, c) = a + bx + cx^2$$

This page titled 5.9: A Isomorphisms is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **9.7: Isomorphisms** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

5.10: The Kernel and Image of a Linear Map

Outcomes

A. Describe the kernel and image of a linear transformation.
B. Use the kernel and image to determine if a linear transformation is one to one or onto.

Here we consider the case where the linear map is not necessarily an isomorphism. First here is a definition of what is meant by the image and kernel of a linear transformation.

Definition 5.10.1: Kernel and Image

Let V and W be vector spaces and let $T : V \to W$ be a linear transformation. Then the image of T denoted as $\text{im}(T)$ is defined to be the set

$$\{T(\vec{v}) : \vec{v} \in V\}$$

In words, it consists of all vectors in W which equal $T(\vec{v})$ for some $\vec{v} \in V$. The kernel, $\ker(T)$, consists of all $\vec{v} \in V$ such that $T(\vec{v}) = \vec{0}$. That is,

$$\ker(T) = \left\{\vec{v} \in V : T(\vec{v}) = \vec{0}\right\}$$

Then in fact, both $\text{im}(T)$ and $\ker(T)$ are subspaces of W and V respectively.

Proposition 5.10.1: Kernel and Image as Subspaces

Let V, W be vector spaces and let $T : V \to W$ be a linear transformation. Then $\ker(T) \subseteq V$ and $\text{im}(T) \subseteq W$. In fact, they are both subspaces.

Proof

First consider $\ker(T)$. It is necessary to show that if $\vec{v}_1, \vec{v}_2$ are vectors in $\ker(T)$ and if a, b are scalars, then $a\vec{v}_1 + b\vec{v}_2$ is also in $\ker(T)$. But

$$T\left(a\vec{v}_1 + b\vec{v}_2\right) = aT(\vec{v}_1) + bT(\vec{v}_2) = a\vec{0} + b\vec{0} = \vec{0}$$

Thus $\ker(T)$ is a subspace of V.

Next suppose $T(\vec{v}_1), T(\vec{v}_2)$ are two vectors in $\text{im}(T)$. Then if a, b are scalars,

$$aT(\vec{v}_2) + bT(\vec{v}_2) = T\left(a\vec{v}_1 + b\vec{v}_2\right)$$

and this last vector is in $\text{im}(T)$ by definition.

Consider the following example.

Example 5.10.1: Kernel and Image of a Transformation

Let $T : \mathbb{P}_1 \to \mathbb{R}$ be the linear transformation defined by

$$T(p(x)) = p(1) \text{ for all } p(x) \in \mathbb{P}_1.$$

Find the kernel and image of T.

Solution

We will first find the kernel of T. It consists of all polynomials in $\mathbb{P}_1$ that have 1 for a root.

$$\begin{aligned}\ker(T) &= \{p(x) \in \mathbb{P}_1 \mid p(1) = 0\} \\ &= \{ax + b \mid a, b \in \mathbb{R} \text{ and } a + b = 0\} \\ &= \{ax - a \mid a \in \mathbb{R}\}\end{aligned}$$

Therefore a basis for $\ker(T)$ is

$$\{x - 1\}$$

Notice that this is a subspace of $\mathbb{P}_1$.

Now consider the image. It consists of all numbers which can be obtained by evaluating all polynomials in $\mathbb{P}_1$ at 1.

$$\begin{aligned}\text{im}(T) &= \{p(1) \mid p(x) \in \mathbb{P}_1\} \\ &= \{a + b \mid ax + b \in \mathbb{P}_1\} \\ &= \{a + b \mid a, b \in \mathbb{R}\} \\ &= \mathbb{R}\end{aligned}$$

Therefore a basis for $\text{im}(T)$ is

$$\{1\}$$

Notice that this is a subspace of $\mathbb{R}$, and in fact is the space $\mathbb{R}$ itself.

Example 5.10.2: Kernel and Image of a Linear Transformation

Let $T : \mathbb{M}_{22} \mapsto \mathbb{R}^2$ be defined by

$$T\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a - b \\ c + d \end{bmatrix}$$

Then T is a linear transformation. Find a basis for $\ker(T)$ and $\text{im}(T)$.

Solution

You can verify that T represents a linear transformation.

Now we want to find a way to describe all matrices A such that $T(A) = \vec{0}$, that is the matrices in $\ker(T)$. Suppose $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is such a matrix. Then

$$T\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a - b \\ c + d \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The values of a, b, c, d that make this true are given by solutions to the system

$$\begin{aligned} a - b &= 0 \\ c + d &= 0 \end{aligned}$$

The solution is $a = s, b = s, c = t, d = -t$ where s, t are scalars. We can describe $\ker(T)$ as follows.

$$\ker(T) = \left\{\begin{bmatrix} s & s \\ t & -t \end{bmatrix}\right\} = \text{span}\left\{\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}\right\}$$

It is clear that this set is linearly independent and therefore forms a basis for $\ker(T)$.

We now wish to find a basis for $\text{im}(T)$. We can write the image of T as

$$\text{im}(T) = \left\{\begin{bmatrix} a - b \\ c + d \end{bmatrix}\right\}$$

Notice that this can be written as

$$\text{span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$$

However this is clearly not linearly independent. By removing vectors from the set to create an independent set gives a basis of $\mathrm{im}(T)$.

$$\left\{\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix}\right\}$$

Notice that these vectors have the same span as the set above but are now linearly independent.

A major result is the relation between the dimension of the kernel and dimension of the image of a linear transformation. A special case was done earlier in the context of matrices. Recall that for an $m\times n$ matrix it was the case that the dimension of the kernel of A added to the rank of A equals n.

Theorem 5.10.1: Dimension of Kernel + Image

Let $T:V\to W$ be a linear transformation where V,W are vector spaces. Suppose the dimension of V is n. Then $n=\dim(\ker(T))+\dim(\mathrm{im}\,(T))$.

Proof

From Proposition 5.10.1, $\mathrm{im}\,(T)$ is a subspace of W. By Theorem 9.4.8, there exists a basis for $\mathrm{im}\,(T)$, $\{T(\vec{v}_1),\cdots,T(\vec{v}_r)\}$. Similarly, there is a basis for $\ker(T)$, $\{\vec{u}_1,\cdots,\vec{u}_s\}$. Then if $\vec{v}\in V$, there exist scalars c_i such that

$$T(\vec{v})=\sum_{i=1}^{r}c_iT(\vec{v}_i)$$

Hence $T\left(\vec{v}-\sum_{i=1}^{r}c_i\vec{v}_i\right)=0$. It follows that $\vec{v}-\sum_{i=1}^{r}c_i\vec{v}_i$ is in $\ker(T)$. Hence there are scalars a_i such that

$$\vec{v}-\sum_{i=1}^{r}c_i\vec{v}_i=\sum_{j=1}^{s}a_j\vec{u}_j$$

Hence $\vec{v}=\sum_{i=1}^{r}c_i\vec{v}_i+\sum_{j=1}^{s}a_j\vec{u}_j$. Since $\vec{v}$ is arbitrary, it follows that

$$V=\mathrm{span}\left\{\vec{u}_1,\cdots,\vec{u}_s,\vec{v}_1,\cdots,\vec{v}_r\right\}$$

If the vectors $\{\vec{u}_1,\cdots,\vec{u}_s,\vec{v}_1,\cdots,\vec{v}_r\}$ are linearly independent, then it will follow that this set is a basis. Suppose then that

$$\sum_{i=1}^{r}c_i\vec{v}_i+\sum_{j=1}^{s}a_j\vec{u}_j=0$$

Apply T to both sides to obtain

$$\sum_{i=1}^{r}c_iT(\vec{v}_i)+\sum_{j=1}^{s}a_jT(\vec{u}_j)=\sum_{i=1}^{r}c_iT(\vec{v}_i)=\vec{0}$$

Since $\{T(\vec{v}_1),\cdots,T(\vec{v}_r)\}$ is linearly independent, it follows that each $c_i=0$. Hence $\sum_{j=1}^{s}a_j\vec{u}_j=0$ and so, since the $\{\vec{u}_1,\cdots,\vec{u}_s\}$ are linearly independent, it follows that each $a_j=0$ also. It follows that $\{\vec{u}_1,\cdots,\vec{u}_s,\vec{v}_1,\cdots,\vec{v}_r\}$ is a basis for V and so

$$n=s+r=\dim(\ker(T))+\dim(\mathrm{im}\,(T))$$

Consider the following definition.

Definition 5.10.2: Rank of Linear Transformation

Let $T:V\to W$ be a linear transformation and suppose V,W are finite dimensional vector spaces. Then the rank of T denoted as $\mathrm{rank}\,(T)$ is defined as the dimension of $\mathrm{im}\,(T)$. The nullity of T is the dimension of $\ker(T)$. Thus the above theorem says that $\mathrm{rank}\,(T)+\dim(\ker(T))=\dim(V)$.

Recall the following important result.

Theorem 5.10.2: Subspace of Same Dimension

Let V be a vector space of dimension n and let W be a subspace. Then $W = V$ if and only if the dimension of W is also n.

From this theorem follows the next corollary.

Corollary 5.10.1: One to One and Onto Characterization

Let $T : V \to W$ be a linear map where the dimension of V is n and the dimension of W is m. Then T is one to one if and only if $\ker(T) = \left\{\vec{0}\right\}$ and T is onto if and only if $\text{rank}\,(T) = m$.

Proof

The statement $\ker(T) = \left\{\vec{0}\right\}$ is equivalent to saying if $T(\vec{v}) = \vec{0}$, it follows that $\vec{v} = \vec{0}$. Thus by Lemma 9.7.1 T is one to one. If T is onto, then $\text{im}\,(T) = W$ and so $\text{rank}\,(T)$ which is defined as the dimension of $\text{im}\,(T)$ is m. If $\text{rank}\,(T) = m$, then by Theorem 5.10.2, since $\text{im}\,(T)$ is a subspace of W, it follows that $\text{im}\,(T) = W$.

Example 5.10.3: One to One Transformation

Let $S : \mathbb{P}_2 \to \mathbb{M}_{22}$ be a linear transformation defined by

$$S(ax^2 + bx + c) = \begin{bmatrix} a+b & a+c \\ b-c & b+c \end{bmatrix} \text{ for all } ax^2 + bx + c \in \mathbb{P}_2.$$

Prove that S is one to one but not onto.

Solution

You may recall this example from earlier in Example 9.7.1. Here we will determine that S is one to one, but not onto, using the method provided in Corollary 5.10.1.

By definition,

$$\ker(S) = \{ax^2 + bx + c \in \mathbb{P}_2 \mid a+b = 0, a+c = 0, b-c = 0, b+c = 0\}.$$

Suppose $p(x) = ax^2 + bx + c \in \ker(S)$. This leads to a homogeneous system of four equations in three variables. Putting the augmented matrix in reduced row-echelon form:

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

Since the unique solution is $a = b = c = 0$, $\ker(S) = \{\vec{0}\}$, and thus S is one-to-one by Corollary 5.10.1.

Similarly, by Corollary 5.10.1, if S is onto it will have $\text{rank}(S) = \dim(\mathbb{M}_{22}) = 4$. The image of S is given by

$$\text{im}(S) = \left\{\begin{bmatrix} a+b & a+c \\ b-c & b+c \end{bmatrix}\right\} = \text{span}\left\{\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}\right\}$$

These matrices are linearly independent which means this set forms a basis for $\text{im}(S)$. Therefore the dimension of $\text{im}(S)$, also called $\text{rank}(S)$, is equal to 3. It follows that S is not onto.

This page titled 5.10: The Kernel and Image of a Linear Map is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **9.8: The Kernel and Image of a Linear Map** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

5.11: The Matrix of a Linear Transformation

Outcomes

A. Find the matrix of a linear transformation with respect to general bases in vector spaces.

You may recall from $\mathbb{R}^n$ that the matrix of a linear transformation depends on the bases chosen. This concept is explored in this section, where the linear transformation now maps from one arbitrary vector space to another.

Let $T : V \mapsto W$ be an isomorphism where V and W are vector spaces. Recall from Lemma 9.7.2 that T maps a basis in V to a basis in W. When discussing this Lemma, we were not specific on what this basis looked like. In this section we will make such a distinction.

Consider now an important definition.

Definition 5.11.1: Coordinate Isomorphism

Let V be a vector space with $\dim(V) = n$, let $B = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n\}$ be a fixed basis of V, and let $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ denote the standard basis of $\mathbb{R}^n$. We define a transformation $C_B : V \to \mathbb{R}^n$ by

$$C_B(a_1\vec{b}_1 + a_2\vec{b}_2 + \cdots + a_n\vec{b}_n) = a_1\vec{e}_1 + a_2\vec{e}_2 + \cdots + a_n\vec{e}_n = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Then C_B is a linear transformation such that $C_B(\vec{b}_i) = \vec{e}_i$, $1 \leq i \leq n$.

C_B is an isomorphism, called the coordinate isomorphism corresponding to B.

We continue with another related definition.

Definition 5.11.2: Coordinate Vector

Let V be a finite dimensional vector space with $\dim(V) = n$, and let $B = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n\}$ be an ordered basis of V (meaning that the order that the vectors are listed is taken into account). The coordinate vector of $\vec{v}$ with respect to B is defined as $C_B(\vec{v})$.

coordinatevector Consider the following example.

Example 5.11.1: Coordinate Vector

Let $V = \mathbb{P}_2$ and $\vec{x} = -x^2 - 2x + 4$. Find $C_B(\vec{x})$ for the following bases B:

1. $B = \{1, x, x^2\}$
2. $B = \{x^2, x, 1\}$
3. $B = \{x + x^2, x, 4\}$

Solution

1. First, note the order of the basis is important. Now we need to find a_1, a_2, a_3 such that $\vec{x} = a_1(1) + a_2(x) + a_3(x^2)$, that is:

$$-x^2 - 2x + 4 = a_1(1) + a_2(x) + a_3(x^2)$$

Clearly the solution is

$$\begin{aligned} a_1 &= 4 \\ a_2 &= -2 \\ a_3 &= -1 \end{aligned}$$

Therefore the coordinate vector is

$$C_B(\vec{x}) = \begin{bmatrix} 4 \\ -2 \\ -1 \end{bmatrix}$$

2. Again remember that the order of B is important. We proceed as above. We need to find a_1, a_2, a_3 such that $\vec{x} = a_1(x^2) + a_2(x) + a_3(1)$, that is:

$$-x^2 - 2x + 4 = a_1(x^2) + a_2(x) + a_3(1)$$

Here the solution is

$$\begin{aligned} a_1 &= -1 \\ a_2 &= -2 \\ a_3 &= 4 \end{aligned}$$

Therefore the coordinate vector is

$$C_B(\vec{x}) = \begin{bmatrix} -1 \\ -2 \\ 4 \end{bmatrix}$$

3. Now we need to find a_1, a_2, a_3 such that $\vec{x} = a_1(x + x^2) + a_2(x) + a_3(4)$, that is:

$$\begin{aligned} -x^2 - 2x + 4 &= a_1(x + x^2) + a_2(x) + a_3(4) \\ &= a_1(x^2) + (a_1 + a_2)(x) + a_3(4) \end{aligned}$$

The solution is

$$\begin{aligned} a_1 &= -1 \\ a_2 &= -1 \\ a_3 &= 1 \end{aligned}$$

and the coordinate vector is

$$C_B(\vec{x}) = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

Given that the coordinate transformation $C_B : V \to \mathbb{R}^n$ is an isomorphism, its inverse exists.

Theorem 5.11.1: Inverse of the Coordinate Isomorphism

Let V be a finite dimensional vector space with dimension n and ordered basis $B = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n\}$. Then $C_B : V \to \mathbb{R}^n$ is an isomorphism whose inverse,

$$C_B^{-1} : \mathbb{R}^n \to V$$

is given by

$$C_B^{-1} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a_1\vec{b}_1 + a_2\vec{b}_2 + \cdots + a_n\vec{b}_n \text{ for all } \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}^n.$$

We now discuss the main result of this section, that is how to represent a linear transformation with respect to different bases.

Let V and W be finite dimensional vector spaces, and suppose

- $\dim(V) = n$ and $B_1 = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n\}$ is an ordered basis of V;
- $\dim(W) = m$ and B_2 is an ordered basis of W.

Let $T : V \to W$ be a linear transformation. If $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$, then we can find a matrix A so that $T_A = T$. For arbitrary vector spaces V and W, our goal is to represent T as a matrix., i.e., find a matrix A so that $T_A : \mathbb{R}^n \to \mathbb{R}^m$ and $T_A = C_{B_2} T C_{B_1}^{-1}$.

To find the matrix A:

$$T_A = C_{B_2} T C_{B_1}^{-1} \text{ implies that } T_A C_{B_1} = C_{B_2} T,$$

and thus for any $\vec{v} \in V$,

$$C_{B_2}[T(\vec{v})] = T_A[C_{B_1}(\vec{v})] = AC_{B_1}(\vec{v}).$$

Since $C_{B_1}(\vec{b}_j) = \vec{e}_j$ for each $\vec{b}_j \in B_1$, $AC_{B_1}(\vec{b}_j) = A\vec{e}_j$, which is simply the jth column of A. Therefore, the jth column of A is equal to $C_{B_2}[T(\vec{b}_j)]$.

The matrix of T corresponding to the ordered bases B_1 and B_2 is denoted $M_{B_2B_1}(T)$ and is given by

$$M_{B_2B_1}(T) = \begin{bmatrix} C_{B_2}[T(\vec{b}_1)] & C_{B_2}[T(\vec{b}_2)] & \cdots & C_{B_2}[T(\vec{b}_n)] \end{bmatrix}.$$

This result is given in the following theorem.

Theorem 5.11.2

Let V and W be vectors spaces of dimension n and m respectively, with $B_1 = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n\}$ an ordered basis of V and B_2 an ordered basis of W. Suppose $T : V \to W$ is a linear transformation. Then the unique matrix $M_{B_2B_1}(T)$ of T corresponding to B_1 and B_2 is given by

$$M_{B_2B_1}(T) = \begin{bmatrix} C_{B_2}[T(\vec{b}_1)] & C_{B_2}[T(\vec{b}_2)] & \cdots & C_{B_2}[T(\vec{b}_n)] \end{bmatrix}.$$

This matrix satisfies $C_{B_2}[T(\vec{v})] = M_{B_2B_1}(T)C_{B_1}(\vec{v})$ for all $\vec{v} \in V$.

We demonstrate this content in the following examples.

Example 5.11.2: Matrix of a Linear Transformation

Let $T : \mathbb{P}_3 \mapsto \mathbb{R}^4$ be an isomorphism defined by

$$T(ax^3 + bx^2 + cx + d) = \begin{bmatrix} a+b \\ b-c \\ c+d \\ d+a \end{bmatrix}$$

Suppose $B_1 = \{x^3, x^2, x, 1\}$ is an ordered basis of $\mathbb{P}_3$ and

$$B_2 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

be an ordered basis of $\mathbb{R}^4$. Find the matrix $M_{B_2B_1}(T)$.

Solution

To find $M_{B_2B_1}(T)$, we use the following definition.

$$M_{B_2B_1}(T) = \begin{bmatrix} C_{B_2}[T(x^3)] & C_{B_2}[T(x^2)] & C_{B_2}[T(x)] & C_{B_2}[T(x^2)] \end{bmatrix}$$

First we find the result of applying T to the basis B_1.

$$T(x^3)=\begin{bmatrix}1\\0\\0\\1\end{bmatrix},T(x^2)=\begin{bmatrix}1\\1\\0\\0\end{bmatrix},T(x)=\begin{bmatrix}0\\-1\\1\\0\end{bmatrix},T(1)=\begin{bmatrix}0\\0\\1\\1\end{bmatrix}$$

Next we apply the coordinate isomorphism C_{B_2} to each of these vectors. We will show the first in detail.

$$C_{B_2}\left(\begin{bmatrix}1\\0\\0\\1\end{bmatrix}\right)=a_1\begin{bmatrix}1\\0\\1\\0\end{bmatrix}+a_2\begin{bmatrix}0\\1\\0\\0\end{bmatrix}+a_3\begin{bmatrix}0\\0\\-1\\0\end{bmatrix}+a_4\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

This implies that

$$\begin{aligned}a_1&=1\\a_2&=0\\a_1-a_3&=0\\a_4&=1\end{aligned}$$

which has a solution given by

$$\begin{aligned}a_1&=1\\a_2&=0\\a_3&=1\\a_4&=1\end{aligned}$$

Therefore $C_{B_2}[T(x^3)]=\begin{bmatrix}1\\0\\1\\1\end{bmatrix}$.

You can verify that the following are true.

$$C_{B_2}[T(x^2)]=\begin{bmatrix}1\\1\\1\\0\end{bmatrix},C_{B_2}[T(x)]=\begin{bmatrix}0\\-1\\-1\\0\end{bmatrix},C_{B_2}[T(1)]=\begin{bmatrix}0\\0\\-1\\1\end{bmatrix}$$

Using these vectors as the columns of $M_{B_2B_1}(T)$ we have

$$M_{B_2B_1}(T)=\begin{bmatrix}1&1&0&0\\0&1&-1&0\\1&1&-1&-1\\1&0&0&1\end{bmatrix}$$

The next example demonstrates that this method can be used to solve different types of problems. We will examine the above example and see if we can work backwards to determine the action of T from the matrix $M_{B_2B_1}(T)$.

Example 5.11.3: Finding the Action of a Linear Transformation

Let $T:\mathbb{P}_3\mapsto\mathbb{R}^4$ be an isomorphism with

$$M_{B_2B_1}(T)=\begin{bmatrix}1&1&0&0\\0&1&-1&0\\1&1&-1&-1\\1&0&0&1\end{bmatrix},$$

where $B_1=\left\{x^3,x^2,x,1\right\}$ is an ordered basis of $\mathbb{P}_3$ and

$$B_2 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is an ordered basis of $\mathbb{R}^4$. If $p(x) = ax^3 + bx^2 + cx + d$, find $T(p(x))$.

Solution

Recall that $C_{B_2}[T(p(x))] = M_{B_2B_1}(T)C_{B_1}(p(x))$. Then we have

$$\begin{aligned} C_{B_2}[T(p(x))] &= M_{B_2B_1}(T)C_{B_1}(p(x)) \\ &= \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & -1 & -1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \\ &= \begin{bmatrix} a+b \\ b-c \\ a+b-c-d \\ a+d \end{bmatrix} \end{aligned}$$

Therefore

$$\begin{aligned} T(p(x)) &= C_D^{-1} \begin{bmatrix} a+b \\ b-c \\ a+b-c-d \\ a+d \end{bmatrix} \\ &= (a+b) \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + (b-c) \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + (a+b-c-d) \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix} + (a+d) \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} a+b \\ b-c \\ c+d \\ a+d \end{bmatrix} \end{aligned}$$

You can verify that this was the definition of $T(p(x))$ given in the previous example.

We can also find the matrix of the composite of multiple transformations.

Theorem 5.11.3: Matrix of Composition

Let V, W and U be finite dimensional vector spaces, and suppose $T : V \mapsto W$, $S : W \mapsto U$ are linear transformations. Suppose V, W and U have ordered bases of B_1, B_2 and B_3 respectively. Then the matrix of the composite transformation $S \circ T$ (or ST) is given by

$$M_{B_3B_1}(ST) = M_{B_3B_2}(S)M_{B_2B_1}(T).$$

The next important theorem gives a condition on when T is an isomorphism.

Theorem 5.11.4: Isomorphism

Let V and W be vector spaces such that both have dimension n and let $T : V \mapsto W$ be a linear transformation. Suppose B_1 is an ordered basis of V and B_2 is an ordered basis of W.

Then the conditions that $M_{B_2B_1}(T)$ is invertible for **all** B_1 and B_2, and that $M_{B_2B_1}(T)$ is invertible for **some** B_1 and B_2 are equivalent. In fact, these occur if and only if T is an isomorphism.

If T is an isomorphism, the matrix $M_{B_2B_1}(T)$ is invertible and its inverse is given by $[M_{B_2B_1}(T)]^{-1} = M_{B_1B_2}(T^{-1})$.

Consider the following example.

Example 5.11.4

Suppose $T : \mathbb{P}_3 \to \mathbb{M}_{22}$ is a linear transformation defined by

$$T(ax^3+bx^2+cx+d) = \begin{bmatrix} a+d & b-c \\ b+c & a-d \end{bmatrix}$$

for all $ax^3+bx^2+cx+d \in \mathbb{P}_3$. Let $B_1 = \{x^3, x^2, x, 1\}$ and

$$B_2 = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

be ordered bases of $\mathbb{P}_3$ and $\mathbb{M}_{22}$, respectively.

1. Find $M_{B_2B_1}(T)$.
2. Verify that T is an isomorphism by proving that $M_{B_2B_1}(T)$ is invertible.
3. Find $M_{B_1B_2}(T^{-1})$, and verify that $M_{B_1B_2}(T^{-1}) = [M_{B_2B_1}(T)]^{-1}$.
4. Use $M_{B_1B_2}(T^{-1})$ to find T^{-1}.

Solution

1. $$\begin{aligned} M_{B_2B_1}(T) &= [\, C_{B_2}[T(1)] \quad C_{B_2}[T(x)] \quad C_{B_2}[T(x^2)] \quad C_{B_2}[T(x^3)]\,] \\ &= \left[C_{B_2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad C_{B_2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad C_{B_2}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad C_{B_2}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right] \\ &= \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \end{aligned}$$

2. $\det(M_{B_2B_1}(T)) = 4$, so the matrix is invertible, and hence T is an isomorphism.

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

   so

   $$T^{-1}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \frac{1+x^3}{2}, T^{-1}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \frac{x-x^2}{2},$$
   $$T^{-1}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \frac{x+x^2}{2}, T^{-1}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \frac{1-x^3}{2}.$$

   Therefore,

   $$M_{B_1B_2}(T^{-1}) = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}$$

   You should verify that $M_{B_2B_1}(T)M_{B_1B_2}(T^{-1}) = I_4$. From this it follows that $[M_{B_2B_1}(T)]^{-1} = M_{B_1B_2}(T^{-1})$.

4.

$$
\begin{aligned}
C_{B_1}\left(T^{-1}\begin{bmatrix} p & q \\ r & s \end{bmatrix}\right) &= M_{B_1B_2}(T^{-1})C_{B_2}\left(\begin{bmatrix} p & q \\ r & s \end{bmatrix}\right) \\
T^{-1}\begin{bmatrix} p & q \\ r & s \end{bmatrix} &= C_{B_1}^{-1}\left(M_{B_1B_2}(T^{-1})C_{B_2}\left(\begin{bmatrix} p & q \\ r & s \end{bmatrix}\right)\right) \\
&= C_{B_1}^{-1}\left(\frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} p \\ q \\ r \\ s \end{bmatrix}\right) \\
&= C_{B_1}^{-1}\left(\frac{1}{2}\begin{bmatrix} p+s \\ q+r \\ r-q \\ p-s \end{bmatrix}\right) \\
&= \frac{1}{2}(p+s)x^3+\frac{1}{2}(q+r)x^2+\frac{1}{2}(r-q)x+\frac{1}{2}(p-s).
\end{aligned}
$$

This page titled 5.11: The Matrix of a Linear Transformation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **9.9: The Matrix of a Linear Transformation** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

5.E: Exercises

Exercise 5.E.1

Suppose you have $\mathbb{R}^2$ and the $+$ operation is as follows:

$$(a,b)+(c,d)=(a+d,b+c).$$

Scalar multiplication is defined in the usual way. Is this a vector space? Explain why or why not.

Exercise 5.E.2

Suppose you have $\mathbb{R}^2$ and the $+$ operation is as follows:

$$(a,b)+(c,d)=(0,b+d)$$

Scalar multiplication is defined in the usual way. Is this a vector space? Explain why or why not.

Exercise 5.E.3

Suppose you have $\mathbb{R}^2$ and scalar multiplication is defined as $c(a,b)=(a,cb)$ while vector addition is defined as usual. Is this a vector space? Explain why or why not.

Exercise 5.E.4

Suppose you have $\mathbb{R}^2$ and the $+$ operation is defined as follows.

$$(a,b)+(c,d)=(a-c,b-d)$$

Scalar multiplication is same as usual. Is this a vector space? Explain why or why not.

Exercise 5.E.5

Consider all the functions defined on a non empty set which have values in $\mathbb{R}$. Is this a vector space? Explain. The operations are defined as follows. Here f,g signify functions and a is a scalar

$$\begin{aligned}(f+g)(x) &= f(x)+g(x)\\ (af)(x) &= a(f(x))\end{aligned}$$

Exercise 5.E.6

Denote by $\mathbb{R}^{\mathbb{N}}$ the set of real valued sequences. For $\vec{a}\equiv\{a_n\}_{n=1}^{\infty}$, $\vec{b}\equiv\{b_n\}_{n=1}^{\infty}$ two of these, define their sum to be given by

$$\vec{a}+\vec{b}=\{a_n+b_n\}_{n=1}^{\infty}$$

and define scalar multiplication by

$$c\vec{a}=\{ca_n\}_{n=1}^{\infty} \text{ where } \vec{a}=\{a+n\}_{n=1}^{\infty}$$

Is this a special case of Exercise 5.E.5? Is this a vector space?

Exercise 5.E.7

Let $\mathbb{C}^2$ be the set of ordered pairs of complex numbers. Define addition and scalar multiplication in the usual way.

$$(z,w)+(\hat{z},\hat{w})=(z+\hat{z},w+\hat{w}),u(z,w)\equiv(uz,uw)$$

Here the scalars are from $\mathbb{C}$. Show this is a vector space.

Exercise 5.E.8

Let V be the set of functions defined on a nonempty set which have values in a vector space W. Is this a vector space? Explain.

Exercise 5.E.9

Consider the space of $m \times n$ matrices with operation of addition and scalar multiplication defined the usual way. That is, if A, B are two $m \times n$ matrices and c a scalar,

$$(A+B)_{ij} = A_{ij} + B_{ij},\ (cA)_{ij} \equiv c(A_{ij})$$

Exercise 5.E.10

Consider the set of $n \times n$ symmetric matrices. That is, $A = A^T$. In other words, $A_{ij} = A_{ji}$. Show that this set of symmetric matrices is a vector space and a subspace of the vector space of $n \times n$ matrices.

Exercise 5.E.11

Consider the set of all vectors in $\mathbb{R}^2, (x, y)$ such that $x + y \geq 0$. Let the vector space operations be the usual ones. Is this a vector space? Is it a subspace of $\mathbb{R}^2$?

Exercise 5.E.12

Consider the vectors in $\mathbb{R}^2, (x, y)$ such that $xy = 0$. Is this a subspace of $\mathbb{R}^2$? Is it a vector space? The addition and scalar multiplication are the usual operations.

Exercise 5.E.13

Define the operation of vector addition on $\mathbb{R}^2$ by $(x, y) + (u, v) = (x + u, y + v + 1)$. Let scalar multiplication be the usual operation. Is this a vector space with these operations? Explain.

Exercise 5.E.14

Let the vectors be real numbers. Define vector space operations in the usual way. That is $x + y$ means to add the two numbers and xy means to multiply them. Is $\mathbb{R}$ with these operations a vector space? Explain.

Exercise 5.E.15

Let the scalars be the rational numbers and let the vectors be real numbers which are the form $a + b\sqrt{2}$ for a, b rational numbers. Show that with the usual operations, this is a vector space.

Exercise 5.E.16

Let $\mathbb{P}_2$ be the set of all polynomials of degree 2 or less. That is, these are of the form $a + bx + cx^2$. Addition is defined as

$$(a + bx + cx^2) + (\hat{d} + \hat{b}x + \hat{c}x^2) = (a + \hat{a}) + (b + \hat{b})x + (c + \hat{c})x^2$$

and scalar multiplication is defined as

$$d(a + bx + cx^2) = da + dbx + cdx^2$$

Show that, with this definition of the vector space operations that $\mathbb{P}_2$ is a vector space. Now let V denote those polynomials $a + bx + cx^2$ such that $a + b + c = 0$. Is V a subspace of $\mathbb{P}_2$? Explain.

Exercise 5.E.17

Let M, N be subspaces of a vector space V and consider $M+N$ defined as the set of all $m+n$ where $m \in M$ and $n \in N$. Show that $M+N$ is a subspace of V.

Exercise 5.E.18

Let M, N be subspaces of a vector space V. Then $M \cap N$ consists of all vectors which are in both M and N. Show that $M \cap N$ is a subspace of V.

Exercise 5.E.19

Let M, N be subspaces of a vector space $\mathbb{R}^2$. Then $N \cup M$ consists of all vectors which are in either M or N. Show that $N \cup M$ is not necessarily a subspace of $\mathbb{R}^2$ by giving an example where $N \cup M$ fails to be a subspace.

Exercise 5.E.20

Let X consist of the real valued functions which are defined on an interval $[a, b]$. For $f, g \in X$, $f+g$ is the name of the function which satisfies $(f+g)(x) = f(x)+g(x)$. For s a real number, $(sf)(x) = s(f(x))$. Show this is a vector space.

Answer

The axioms of a vector space all hold because they hold for a vector space. The only thing left to verify is the assertions about the things which are supposed to exist. 0 would be the zero function which sends everything to 0. This is an additive identity. Now if f is a function, $-f(x) \equiv (-f(x))$. Then

$$(f+(-f))(x) \equiv f(x)+(-f)(x) \equiv f(x)+(-f(x)) = 0$$

Hence $f+-f = 0$. For each $x \in [a, b]$, let $f_x(x) = 1$ and $f_x(y) = 0$ if $y \neq x$. Then these vectors are obviously linearly independent.

Exercise 5.E.21

Consider functions defined on $\{1, 2, \cdots, n\}$ having values in $\mathbb{R}$. Explain how, if V is the set of all such functions, V can be considered as $\mathbb{R}^n$.

Answer

Let $f(i)$ be the ith component of a vector $\vec{x} \in \mathbb{R}^n$. Thus a typical element in $\mathbb{R}^n$ is $(f(1), \cdots, f(n))$.

Exercise 5.E.22

Let the vectors be polynomials of degree no more than 3. Show that with the usual definitions of scalar multiplication and addition wherein, for $p(x)$ a polynomial, $(ap)(x) = ap(x)$ and for p, q polynomials $(p+q)(x) = p(x)+q(x)$, this is a vector space.

Answer

This is just a subspace of the vector space of functions because it is closed with respect to vector addition and scalar multiplication. Hence this is a vector space.

Exercise 5.E. 23

Let V be a vector space and suppose $\{\vec{x}_1, \cdots, \vec{x}_l\}$ is a set of vectors in V. Show that $\vec{0}$ is in $span\{\vec{x}_1, \cdots, \vec{x}_k\}$.

Answer

$$\sum_{i=1}^{k} 0\vec{x}_k = \vec{0}$$

Exercise 5.E. 24

Determine if $p(x) = 4x^2 - x$ is in the span given by

$$span\{x^2 + x,\ x^2 - 1,\ -x + 2\}$$

Exercise 5.E. 25

Determine if $p(x) = -x^2 + x + 2$ is in the span given by

$$span\{x^2 + x + 1,\ 2x^2 + x\}$$

Exercise 5.E. 26

Determine if $A = \begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix}$ is in the span given by

$$span\left\{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}\right\}$$

Exercise 5.E. 27

Show that the spanning set in Exercise 5.E. 26 is a spanning set for M_{22}, the vector space of all 2×2 matrices.

Exercise 5.E. 28

Consider the vector space of polynomials of degree at most 2, $\mathbb{P}_2$. Determine whether the following is a basis for $\mathbb{P}_2$.

$$\{x^2 + x + 1,\ 2x^2 + 2x + 1,\ x + 1\}$$

Hint: There is a isomorphism from $\mathbb{R}^3$ to $\mathbb{P}_2$. It is defined as follows:

$$T\vec{e}_1 = 1,\ T\vec{e}_2 = x,\ T\vec{e}_3 = x^2$$

Then extend T linearly. Thus

$$T\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = x^2 + x + 1,\ T\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = 2x^2 + 2x + 1,\ T\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 1 + x$$

It follows that if

$$\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right\}$$

is a basis for $\mathbb{R}^3$, then the polynomials will be a basis for $\mathbb{P}_2$ because they will be independent. Recall that an isomorphism takes a linearly independent set to a linearly independent set. Also, since T is an isomorphism, it preserves all linear relations.

Exercise 5.E.29

Find a basis in $\mathbb{P}_2$ for the subspace

$$span\{1+x+x^2,\ 1+2x,\ 1+5x-3x^2\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others. Hint: This is the situation in which you have a spanning set and you want to cut it down to form a linearly independent set which is also a spanning set. Use the same isomorphism above. Since T is an isomorphism, it preserves all linear relations so if such can be found in $\mathbb{R}^3$, the same linear relations will be present in $\mathbb{P}_2$.

Exercise 5.E.30

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{1+x-x^2+x^3,\ 1+2x+3x^3,\ -1+3x+5x^2+7x^3,\ 1+6x+4x^2+11x^3\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.31

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{1+x-x^2+x^3,\ 1+2x+3x^3,\ -1+3x+5x^2+7x^3,\ 1+6x+4x^2+11x^3\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.32

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{x^3-2x^2+x+2,\ 3x^3-x^2+2x+2,\ 7x^3+x^2+4x+2,\ 5x^3+3x+2\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.33

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{x^3+2x^2+x-2,\ 3x^3+3x^2+2x-2,\ 3x^3+x+2,\ 3x^3+x+2\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.34

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{x^3-5x^2+x+5,\ 3x^3-4x^2+2x+5,\ 5x^3+8x^2+2x-5,\ 11x^3+6x+5\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.35

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{x^3-3x^2+x+3,\ 3x^3-2x^2+2x+3,\ 7x^3+7x^2+3x-3,\ 7x^3+4x+3\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.36

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{x^3-x^2+x+1,\ 3x^3+2x+1,\ 4x^3+x^2+2x+1,\ 3x^3+2x-1\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.37

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{x^3-x^2+x+1,\ 3x^3+2x+1,\ 13x^3+x^2+8x+4,\ 3x^3+2x-1\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.38

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{x^3-3x^2+x+3,\ 3x^3-2x^2+2x+3,\ -5x^3+5x^2-4x-6,\ 7x^3+4x-3\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.39

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{x^3-2x^2+x+2,\ 3x^3-x^2+2x+2,\ 7x^3-x^2+4x+4,\ 5x^3+3x-2\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.40

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{x^3-2x^2+x+2,\ 3x^3-x^2+2x+2,\ 3x^3+4x^2+x-2,\ 7x^3-x^2+4x+4\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.41

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{x^3-4x^2+x+4,\ 3x^3-3x^2+2x+4,\ -3x^3+3x^2-2x-4,\ -2x^3+4x^2-2x-4\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.42

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{x^3+2x^2+x-2,\ 3x^3+3x^2+2x-2,\ 5x^3+x^2+2x+2,\ 10x^3+10x^2+6x-6\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.43

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{x^3+x^2+x-1,\ 3x^3+2x^2+2x-1,\ x^3+1,\ 4x^3+3x^2+2x-1\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.44

Find a basis in $\mathbb{P}_3$ for the subspace

$$span\{x^3-x^2+x+1,\ 3x^3+2x+1,\ x^3+2x^2-1,\ 4x^3+x^2+2x+1\}$$

If the above three vectors do not yield a basis, exhibit one of them as a linear combination of the others.

Exercise 5.E.45

Here are some vectors.

$$\{x^3+x^2-x-1,\ 3x^3+2x^2+2x-1\}$$

If these are linearly independent, extend to a basis for all of $\mathbb{P}_3$.

Exercise 5.E.46

Here are some vectors.

$$\{x^3-2x^2-x+2,\ 3x^3-x^2+2x+2\}$$

If these are linearly independent, extend to a basis for all of $\mathbb{P}_3$.

Exercise 5.E.47

Here are some vectors.

$$\{x^3-3x^2-x+3,\ 3x^3-2x^2+2x+3\}$$

If these are linearly independent, extend to a basis for all of $\mathbb{P}_3$.

Exercise 5.E.48

Here are some vectors.

$$\{x^3-2x^2-3x+2,\ 3x^3-x^2-6x+2,\ -8x^3+18x+10\}$$

If these are linearly independent, extend to a basis for all of $\mathbb{P}_3$.

Exercise 5.E.49

Here are some vectors.

$$\{x^3-3x^2-3x+3,\ 3x^3-2x^2-6x+3,\ -8x^3+18x+40\}$$

If these are linearly independent, extend to a basis for all of $\mathbb{P}_3$.

Exercise 5.E.50

Here are some vectors.

$$\{x^3 - x^2 + x + 1,\ 3x^3 + 2x + 1,\ 4x^3 + 2x + 2\}$$

If these are linearly independent, extend to a basis for all of $\mathbb{P}_3$.

Exercise 5.E.51

Here are some vectors.

$$\{x^3 + x^2 + 2x - 1,\ 3x^3 + 2x^2 + 4x - 1,\ 7x^3 + 8x + 23\}$$

If these are linearly independent, extend to a basis for all of $\mathbb{P}_3$.

Exercise 5.E.52

Determine if the following set is linearly independent. If it is linearly dependent, write one vector as a linear combination of the other vectors in the set.

$$\{x + 1,\ x^2 + 2,\ x^2 - x - 3\}$$

Exercise 5.E.53

Determine if the following set is linearly independent. If it is linearly dependent, write one vector as a linear combination of the other vectors in the set.

$$\{x^2 + x,\ -2x^2 - 4x - 6,\ 2x - 2\}$$

Exercise 5.E.54

Determine if the following set is linearly independent. If it is linearly dependent, write one vector as a linear combination of the other vectors in the set.

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

Exercise 5.E.55

Determine if the following set is linearly independent. If it is linearly dependent, write one vector as a linear combination of the other vectors in the set.

$$\left\{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}\right\}$$

Exercise 5.E.56

If you have 5 vectors in $\mathbb{R}^5$ and the vectors are linearly independent, can it always be concluded they span $\mathbb{R}^5$?

Answer

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.

Exercise 5.E.57

If you have 6 vectors in $\mathbb{R}^5$, is it possible they are linearly independent? Explain.

Answer

No. They can't be.

Exercise 5.E.58

Let $\mathbb{P}_3$ be the polynomials of degree no more than 3. Determine which of the following are bases for this vector space.

a. $\{x+1,\ x^3+x^2+2x,\ x^2+x,\ x^3+x^2+x\}$
b. $\{x^3+1,\ x^2+x,\ 2x^3+x^2,\ 2x^3-x^2-3x+1\}$

Answer

a.

b. Suppose

$$c_1(x^3+1)+c_2(x^2+x)+c_3(2x^3+x^2)+c_4(2x^3-x^2-3x+1)=0$$

Then combine the terms according to power of x.

$$(c_1+2c_3+2c_4)x^3+(c_2+c_3-c_4)x^2+(c_2-3c_4)x+(c_1+c_4)=0$$

Is there a non zero solution to the system

$$\begin{aligned} c_1+2c_3+2c_4 &= 0\\ c_2+c_3-c_4 &= 0\\ c_2-3c_4 &= 0\\ c_1+c_4 &= 0 \end{aligned}$$

, Solution is:

$$[c_1=0,\ c_2=0,\ c_3=0,\ c_4=0]$$

Therefore, these are linearly independent.

Exercise 5.E.59

In the context of the above problem, consider polynomials

$$\{a_ix^3+b_ix^2+c_ix+d_i,\ i=1,2,3,4\}$$

Show that this collection of polynomials is linearly independent on an interval $[s,t]$ if and only if

$$\begin{bmatrix} a_1 & b_1 & c_1 & d_1\\ a_2 & b_2 & c_2 & d_2\\ a_3 & b_3 & c_3 & d_3\\ a_4 & b_4 & c_4 & d_4 \end{bmatrix}$$

is an invertible matrix.

Answer

Let $p_i(x)$ denote the ith of these polynomials. Suppose $\sum_i C_ip_i(x)=0$. Then collecting terms according to the exponent of x, you need to have

$$\begin{aligned} C_1a_1 + C_2a_2 + C_3a_3 + C_4a_4 &= 0 \\ C_1b_1 + C_2b_2 + C_3b_3 + C_4b_4 &= 0 \\ C_1c_1 + C_2c_2 + C_3c_3 + C_4c_4 &= 0 \\ C_1d_1 + C_2d_2 + C_3d_3 + C_4d_4 &= 0 \end{aligned}$$

The matrix of coefficients is just the transpose of the above matrix. There exists a non trivial solution if and only if the determinant of this matrix equals 0.

? Exercise 5.E.60

Let the field of scalars be $\mathbb{Q}$, the rational numbers and let the vectors be of the form $a+b\sqrt{2}$ where a, b are rational numbers. Show that this collection of vectors is a vector space with field of scalars $\mathbb{Q}$ and give a basis for this vector space.

Answer

When you add two of these you get one and when you multiply one of these by a scalar, you get another one. A basis is $\{1, \sqrt{2}\}$. By definition, the span of these gives the collection of vectors. Are they independent? Say $a+b\sqrt{2}=0$ where a, b are rational numbers. If $a \neq 0$, then $b\sqrt{2} = -a$ which can't happen since a is rational. If $b \neq 0$, then $-a = b\sqrt{2}$ which again can't happen because on the left is a rational number and on the right is an irrational. Hence both $a, b = 0$ and so this is a basis.

? Exercise 5.E.61

Suppose V is a finite dimensional vector space. Based on the exchange theorem above, it was shown that any two bases have the same number of vectors in them. Give a different proof of this fact using the earlier material in the book. Hint: Suppose $\{\vec{x}_1, \cdots, \vec{x}_n\}$ and $\{\vec{y}_1, \cdots, \vec{y}_m\}$ are two bases with $m < n$. Then define

$$\varphi : \mathbb{R}^n \mapsto V, \ \psi : \mathbb{R}^m \mapsto V$$

by

$$\varphi(\vec{a}) = \sum_{k=1}^{n} a_k \vec{x}_k, \ \psi(\vec{b}) = \sum_{j=1}^{m} b_j \vec{y}_j$$

Consider the linear transformation, $\psi^{-1} \circ \varphi$. Argue it is a one to one and onto mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$. Now consider a matrix of this linear transformation and its reduced row-echelon form.

Answer

This is obvious because when you add two of these you get one and when you multiply one of these by a scalar, you get another one. A basis is $\{1, \sqrt{2}\}$. By definition, the span of these gives the collection of vectors. Are they independent? Say $a+b\sqrt{2}=0$ where a, b are rational numbers. If $a \neq 0$, then $b\sqrt{2} = -a$ which can't happen since a is rational. If $b \neq 0$, then $-a = b\sqrt{2}$ which again can't happen because on the left is a rational number and on the right is an irrational. Hence both $a, b = 0$ and so this is a basis.

? Exercise 5.E.62

Let $M = \{\vec{u} = (u_1, \ u_2, \ u_3, \ u_4) \in \mathbb{R}^4 \ : \ |u_1| \leq 4\}$. Is M a subspace of $\mathbb{R}^4$?

Answer

This is not a subspace. $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ is in it, but $20 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ is not.

Exercise 5.E.63

Let $M = \{\vec{u} = (u_1,\ u_2,\ u_3,\ u_4) \in \mathbb{R}^4\ :\ \sin(u_1) = 1\}$. Is M a subspace of $\mathbb{R}^4$?

Answer

This is not a subspace.

Exercise 5.E.64

Let W be a subset of M_{22} given by

$$W = \{A | A \in M_{22}, A^T = A\}$$

In words, W is the set of all symmetric 2×2 matrices. Is W a subspace of M_{22}?

Exercise 5.E.65

Let W be a subset of M_{22} given by

$$W = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \ | a, b, c, d \in \mathbb{R},\ a + b = c + d \right\}$$

Is W a subspace of M_{22}?

Exercise 5.E.66

Let W be a subset of P_3 given by

$$W = \{ax^3 + bx^2 + cx + d |\ a, b, c, d \in \mathbb{R}, d = 0\}$$

Is W a subspace of P_3?

Exercise 5.E.67

Let W be a subset of P_3 given by

$$W = \{p(x) = ax^3 + bx^2 + cx + d |\ a, b, c, d \in \mathbb{R}, p(2) = 1\}$$

Is W a subspace of P_3?

Exercise 5.E.68

Let T: $\mathbb{P}_2 \to \mathbb{R}$ be a linear transformation such that

$$T(x^2) = 1;\ T(x^2 + x) = 5;\ T(x^2 + x + 1) = -1.$$

Find $T(ax^2 + bx + c)$.

Answer

By linearity we have $T(x^2) = 1$, $T(x) = T(x^2 + x - x^2) = T(x^2 + x) - T(x^2) = 5 - 1 = 5$, and $T(1) = T(x^2 + x + 1 - (x^2 + x)) = T(x^2 + x + 1) - T(x^2 + x)) = -1 - 5 = -6$. Thus $T(ax^2 + bx + c) = aT(x^2) + bT(x) + cT(1) = a + 5b - 6c$.

https://math.libretexts.org/@go/page/170476

Exercise 5.E.69

Consider the following functions T: $\mathbb{R}^3 \to \mathbb{R}^2$. Explain why each of these functions T is not linear.

a. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+2y+3z+1 \\ 2y-3x+z \end{bmatrix}$

b. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+2y^2+3z \\ 2y+3z+z \end{bmatrix}$

c. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \sin x+2y+3z \\ 2y+3z+z \end{bmatrix}$

d. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+2y+3z \\ 2y+3z-\ln z \end{bmatrix}$

Exercise 5.E.70

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 $T(\vec{x}) = A\vec{x}$.

Answer

$$\begin{bmatrix} 3 & 1 & 1 \\ 3 & 2 & 3 \\ 3 & 3 & -1 \end{bmatrix}\begin{bmatrix} 6 & 2 & 1 \\ 5 & 2 & 1 \\ 6 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 29 & 9 & 5 \\ 46 & 13 & 8 \\ 27 & 11 & 5 \end{bmatrix}$$

Exercise 5.E.71

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 $T(\vec{x}) = A\vec{x}$.

Answer

$$\begin{bmatrix} 5 & 3 & 2 \\ 2 & 3 & 5 \\ 5 & 5 & -2 \end{bmatrix} \begin{bmatrix} 11 & 4 & 1 \\ 10 & 4 & 1 \\ 12 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 109 & 38 & 10 \\ 112 & 35 & 10 \\ 81 & 34 & 8 \end{bmatrix}$$

Exercise $5.E.72$

Consider the following functions $T\colon \mathbb{R}^3 \to \mathbb{R}^2$. Show that each is a linear transformation and determine for each the matrix A such that $T(\vec{x}) = A\vec{x}$.

a. $T \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+2y+3z \\ 2y-3x+z \end{bmatrix}$

b. $T \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 7x+2y+z \\ 3x-11y+2z \end{bmatrix}$

c. $T \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3x+2y+z \\ x+2y+6z \end{bmatrix}$

d. $T \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2y-5x+z \\ x+y+z \end{bmatrix}$

Exercise $5.E.73$

Suppose

$$[A_1 \cdots A_n]^{-1}$$

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 $5.E.74$

Let V and W be subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively and let $T\colon V \to W$ be a linear transformation. Suppose that $\{T\vec{v}_1, \cdots, T\vec{v}_r\}$ is linearly independent. Show that it must be the case that $\{\vec{v}_1, \cdots, \vec{v}_r\}$ is also linearly independent.

Answer

If $\sum_i^r a_i \vec{v}_r = 0$, then using linearity properties of T we get

$$0 = T(0) = T\left(\sum_i^r a_i \vec{v}_r\right) = \sum_i^r a_i T(\vec{v}_r).$$

Since we assume that $\{T\vec{v}_1, \cdots, T\vec{v}_r\}$ is linearly independent, we must have all $a_i = 0$, and therefore we conclude that $\{\vec{v}_1, \cdots, \vec{v}_r\}$ is also linearly independent.

Exercise $5.E.75$

Let

https://math.libretexts.org/@go/page/170476

$$V = span\left\{\begin{bmatrix}1\\1\\2\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\0\\1\end{bmatrix}\right\}$$

Let $T\vec{x} = A\vec{x}$ where A is the matrix

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

Give a basis for $Im(T)$.

? Exercise $5.E.76$

Let

$$V = span\left\{\begin{bmatrix}1\\0\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\4\\4\\1\end{bmatrix}\right\}$$

Let $T\vec{x} = A\vec{x}$ where A is the matrix

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

Find a basis for $Im(T)$. In this case, the original vectors do not form an independent set.

Answer

Since the third vector is a linear combinations of the first two, then the image of the third vector will also be a linear combinations of the image of the first two. However the image of the first two vectors are linearly independent (check!), and hence form a basis of the image. Thus a basis for $Im(T)$ is:

$$V = span\left\{\begin{bmatrix}2\\0\\1\\3\end{bmatrix}, \begin{bmatrix}4\\2\\4\\5\end{bmatrix}\right\}$$

? Exercise $5.E.77$

If $\{\vec{v}_1, \cdots, \vec{v}_r\}$ is linearly independent and T is a one to one linear transformation, show that $\{T\vec{v}_1, \cdots, T\vec{v}_r\}$ is also linearly independent. Give an example which shows that if T is only linear, it can happen that, although $\{\vec{v}_1, \cdots, \vec{v}_r\}$ is linearly independent, $\{T\vec{v}_1, \cdots, T\vec{v}_r\}$ is not. In fact, show that it can happen that each of the $T\vec{v}_j$ equals 0.

? Exercise $5.E.78$

Let V and W be subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively and let $T: V \to W$ be a linear transformation. Show that if T is onto W and if $\{\vec{v}_1, \cdots, \vec{v}_r\}$ is a basis for V, then $span\{T\vec{v}_1, \cdots, T\vec{v}_r\} = W$.

https://math.libretexts.org/@go/page/170476

Exercise 5.E.79

Define $T: \mathbb{R}^4 \to \mathbb{R}^3$ as follows.

$$T\vec{x} = \begin{bmatrix} 3 & 2 & 1 & 8 \\ 2 & 2 & -2 & 6 \\ 1 & 1 & -1 & 3 \end{bmatrix} \vec{x}$$

Find a basis for $Im(T)$. Also find a basis for $\ker(T)$.

Exercise 5.E.80

Define $T: \mathbb{R}^4 \to \mathbb{R}^3$ as follows.

$$T\vec{x} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \vec{x}$$

where on the right, it is just matrix multiplication of the vector $\vec{x}$ which is meant. Explain why T is an isomorphism of $\mathbb{R}^3$ to $\mathbb{R}^3$.

Exercise 5.E.81

Suppose $T: \mathbb{R}^3 \to \mathbb{R}^3$ is a linear transformation given by

$$T\vec{x} = A\vec{x}$$

where A is a 3×3 matrix. Show that T is an isomorphism if and only if A is invertible.

Exercise 5.E.82

Suppose $T: \mathbb{R}^3 \to \mathbb{R}^3$ is a linear transformation given by

$$T\vec{x} = A\vec{x}$$

where A is a $m \times n$ matrix. Show that T is never an isomorphism if $m \neq n$. In particular, show that if $m > n$, T cannot be onto and if $m < n$, then T cannot be one to one.

Exercise 5.E.83

Define $T: \mathbb{R}^2 \to \mathbb{R}^3$ as follows.

$$T\vec{x} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \vec{x}$$

where on the right, it is just matrix multiplication of the vector $\vec{x}$ which is meant. Show that T is one to one. Next let $W = Im(T)$. Show that T is an isomorphism of $\mathbb{R}^2$ and $Im(T)$.

Exercise 5.E.84

In the above problem, find a 2×3 matrix A such that the restriction of A to $Im(T)$ gives the same result as T^{-1} on $Im(T)$. Hint: You might let A be such that

$$A \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \; A \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

now find another vector $\vec{v} \in \mathbb{R}^3$ such that

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \vec{v} \right\}$$

is a basis. You could pick

$$\vec{v} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

for example. Explain why this one works or one of your choice works. Then you could define $A\vec{v}$ to equal some vector in $\mathbb{R}^2$. Explain why there will be more than one such matrix A which will deliver the inverse isomorphism T^{-1} on $Im(T)$.

? Exercise $5.E.85$

Now let V equal $span \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$ and let $T: V \to W$ be a linear transformation where

$$W = span \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

and

$$T \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, T \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Explain why T is an isomorphism. Determine a matrix A which, when multiplied on the left gives the same result as T on V and a matrix B which delivers T^{-1} on W. Hint: You need to have

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

Now enlarge $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ to obtain a basis for $\mathbb{R}^3$. You could add in $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ for example, and then pick another vector in $\mathbb{R}^4$ and let $A \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ equal this other vector. Then you would have

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

This would involve picking for the new vector in $\mathbb{R}^4$ the vector $\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T$. Then you could find A. You can do something similar to find a matrix for T^{-1} denoted as B.

Exercise 5.E.86

Let $V = \mathbb{R}^3$ and let

$$W = span(S), \text{ where } S = \left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} \right\}$$

Find a basis of W consisting of vectors in S.

Answer

In this case $\dim(W) = 1$ and a basis for W consisting of vectors in S can be obtained by taking any (nonzero) vector from S.

Exercise 5.E.87

Let T be a linear transformation given by

$$T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Find a basis for $\ker(T)$ and $Im(T)$.

Answer

A basis for $\ker(T)$ is $\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ and a basis for $Im(T)$ is $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$. There are many other possibilities for the specific bases, but in this case $\dim(\ker(T)) = 1$ and $\dim(Im(T)) = 1$.

Exercise 5.E.88

Let T be a linear transformation given by

$$T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Find a basis for $\ker(T)$ and $Im(T)$.

Answer

In this case $\ker(T) = \{0\}$ and $Im(T) = \mathbb{R}^2$ (pick any basis of $\mathbb{R}^2$).

Exercise 5.E.89

Let $V = \mathbb{R}^3$ and let

$$W = span \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} \right\}$$

Extend this basis of W to a basis of V.

Answer

There are many possible such extensions, one is (how do we know?):

$$\left\{\begin{bmatrix}1\\1\\1\end{bmatrix},\begin{bmatrix}-1\\2\\-1\end{bmatrix},\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$$

Exercise 5.E.90

Let T be a linear transformation given by

$$T\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}1&1&1\\1&1&1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}$$

What is $\dim(\ker(T))$?

Answer

We can easily see that $\dim(Im(T))=1$, and thus $\dim(\ker(T))=3-\dim(Im(T))=3-1=2$.

Exercise 5.E.91

Consider the following functions which map $\mathbb{R}^n$ to $\mathbb{R}^n$.

a. T multiplies the jth component of $\vec{x}$ by a nonzero number b.
b. T replaces the ith component of $\vec{x}$ with b times the jth component added to the ith component.
c. T switches the ith and jth components.

Show these functions are linear transformations and describe their matrices A such that $T(\vec{x})=A\vec{x}$.

Answer

a. The matrix of T is the elementary matrix which multiplies the jth diagonal entry of the identity matrix by b.
b. The matrix of T is the elementary matrix which takes b times the jth row and adds to the ith row.
c. The matrix of T is the elementary matrix which switches the ith and the jth rows where the two components are in the ith and jth positions.

Exercise 5.E.92

You are given a linear transformation $T\colon \mathbb{R}^n\to\mathbb{R}^m$ 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

$$\begin{bmatrix}B_1&\cdots&B_n\end{bmatrix}\begin{bmatrix}A_1&\cdots&A_n\end{bmatrix}^{-1}$$

Answer

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

$$\begin{aligned}\begin{bmatrix}\vec{b}_1 & \cdots & \vec{b}_n\end{bmatrix}\begin{bmatrix}\vec{a}_1 & \cdots & \vec{a}_n\end{bmatrix}^{-1}\vec{a}_i &= \begin{bmatrix}\vec{b}_1 & \cdots & \vec{b}_n\end{bmatrix}\begin{bmatrix}\vec{c}_1^{\,T}\\ \vdots \\ \vec{c}_n^{\,T}\end{bmatrix}\vec{a}_i \\ &= \begin{bmatrix}\vec{b}_1 & \cdots & \vec{b}_n\end{bmatrix}\vec{e}_i \\ &= \vec{b}_i\end{aligned}$$

Thus $T\vec{a}_i = \begin{bmatrix}\vec{b}_1 & \cdots & \vec{b}_n\end{bmatrix}\begin{bmatrix}\vec{a}_1 & \cdots & \vec{a}_n\end{bmatrix}^{-1}\vec{a}_i = A\vec{a}_i$. If $\vec{x}$ is arbitrary, then since the matrix $\begin{bmatrix}\vec{a}_1 & \cdots & \vec{a}_n\end{bmatrix}$ is invertible, there exists a unique $\vec{y}$ such that $\begin{bmatrix}\vec{a}_1 & \cdots & \vec{a}_n\end{bmatrix}\vec{y} = \vec{x}$ Hence

$$T\vec{x} = T\left(\sum_{i=1}^{n} y_i\vec{a}_i\right) = \sum_{i=1}^{n} y_i T\vec{a}_i = \sum_{i=1}^{n} y_1 A\vec{a}_i = A\left(\sum_{i=1}^{n} y_i\vec{a}_i\right) = A\vec{x}$$

Exercise $5.E.93$

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 $T(\vec{x}) = A\vec{x}$.

Answer

$$\begin{bmatrix}5 & 1 & 5\\1 & 1 & 3\\3 & 5 & -2\end{bmatrix}\begin{bmatrix}3 & 2 & 1\\2 & 2 & 1\\4 & 1 & 1\end{bmatrix} = \begin{bmatrix}37 & 17 & 11\\17 & 7 & 5\\11 & 14 & 6\end{bmatrix}$$

Exercise $5.E.94$

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 $T(\vec{x}) = A\vec{x}$.

Answer

$$\begin{bmatrix} 1 & 2 & 6 \\ 3 & 4 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 6 & 3 & 1 \\ 5 & 3 & 1 \\ 6 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 52 & 21 & 9 \\ 44 & 23 & 8 \\ 5 & 4 & 1 \end{bmatrix}$$

Exercise $5.E.95$

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 $T(\vec{x}) = A\vec{x}$.

Answer

$$\begin{bmatrix} -3 & 1 & 5 \\ 1 & 3 & 3 \\ 3 & -3 & -3 \end{bmatrix} \begin{bmatrix} 2 & 2 & 1 \\ 1 & 2 & 1 \\ 4 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 15 & 1 & 3 \\ 17 & 11 & 7 \\ -9 & -3 & -3 \end{bmatrix}$$

Exercise $5.E.96$

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 $T(\vec{x}) = A\vec{x}$.

Answer

$$\begin{bmatrix} 3 & 1 & 1 \\ 3 & 2 & 3 \\ 3 & 3 & -1 \end{bmatrix} \begin{bmatrix} 6 & 2 & 1 \\ 5 & 2 & 1 \\ 6 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 29 & 9 & 5 \\ 46 & 13 & 8 \\ 27 & 11 & 5 \end{bmatrix}$$

Exercise $5.E.97$

Suppose T is a linear transformation such that

https://math.libretexts.org/@go/page/170476

$$T\begin{bmatrix}1\\2\\-18\end{bmatrix}=\begin{bmatrix}5\\2\\5\end{bmatrix}$$
$$T\begin{bmatrix}-1\\-1\\15\end{bmatrix}=\begin{bmatrix}3\\3\\5\end{bmatrix}$$
$$T\begin{bmatrix}0\\-1\\4\end{bmatrix}=\begin{bmatrix}2\\5\\-2\end{bmatrix}$$

Find the matrix of T. That is find A such that $T(\vec{x})=A\vec{x}$.

Answer

$$\begin{bmatrix}5&3&2\\2&3&5\\5&5&-2\end{bmatrix}\begin{bmatrix}11&4&1\\10&4&1\\12&3&1\end{bmatrix}=\begin{bmatrix}109&38&10\\112&35&10\\81&34&8\end{bmatrix}$$

Exercise $5.E.98$

Consider the following functions $T\colon \mathbb{R}^3\to\mathbb{R}^2$. Show that each is a linear transformation and determine for each the matrix A such that $T(\vec{x})=A\vec{x}$.

a. $T\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}x+2y+3z\\2y-3x+z\end{bmatrix}$

b. $T\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}7x+2y+z\\3x-11y+2z\end{bmatrix}$

c. $T\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}3x+2y+z\\x+2y+6z\end{bmatrix}$

d. $T\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}2y-5x+z\\x+y+z\end{bmatrix}$

Exercise $5.E.99$

Consider the following functions $T\colon \mathbb{R}^3\to\mathbb{R}^2$. Explain why each of these functions T is not linear.

a. $T\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}x+2y+3z+1\\2y-3x+z\end{bmatrix}$

b. $T\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}x+2y^2+3z\\2y+3x+z\end{bmatrix}$

c. $T\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}\sin x+2y+3z\\2y+3x+z\end{bmatrix}$

d. $T\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}x+2y+3z\\2y+3x-\ln z\end{bmatrix}$

Exercise 5.E.100

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.E.101

Find the matrix for $T(\vec{w}) = \text{proj}_{\vec{v}}(\vec{w})$ where $\vec{v} = \begin{bmatrix} 1 & -2 & 3 \end{bmatrix}^T$.

Answer

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 ith column equal to $\text{proj}_{\vec{u}}(\vec{e}_i)$. Therefore, the matrix desired is

$$\frac{1}{14} \begin{bmatrix} 1 & -2 & 3 \\ -2 & 4 & -6 \\ 3 & -6 & 9 \end{bmatrix}$$

Exercise 5.E.102

Find the matrix for $T(\vec{w}) = \text{proj}_{\vec{v}}(\vec{w})$ where $\vec{v} = \begin{bmatrix} 1 & 5 & 3 \end{bmatrix}^T$.

Answer

$$\frac{1}{35} \begin{bmatrix} 1 & 5 & 3 \\ 5 & 25 & 15 \\ 3 & 15 & 9 \end{bmatrix}$$

Exercise 5.E.103

Find the matrix for $T(\vec{w}) = \text{proj}_{\vec{v}}(\vec{w})$ where $\vec{v} = \begin{bmatrix} 1 & 0 & 3 \end{bmatrix}^T$.

Answer

$$\frac{1}{10} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 0 & 0 \\ 3 & 0 & 9 \end{bmatrix}$$

Exercise 5.E.104

Let $B = \left\{ \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix} \right\}$ be a basis of $\mathbb{R}^2$ and let $\vec{x} = \begin{bmatrix} 5 \\ -7 \end{bmatrix}$ be a vector in $\mathbb{R}^2$. Find $C_B(\vec{x})$.

Exercise 5.E.105

Let $B = \left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \right\}$ be a basis of $\mathbb{R}^3$ and let $\vec{x} = \begin{bmatrix} 5 \\ -1 \\ 4 \end{bmatrix}$ be a vector in $\mathbb{R}^2$. Find $C_B(\vec{x})$.

Answer

$$C_B(\vec{x}) = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}.$$

? Exercise 5.E. 106

Let T: $\mathbb{R}^2 \mapsto \mathbb{R}^2$ be a linear transformation defined by $T\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} a+b \\ a-b \end{bmatrix}$.

Consider the two bases

$$B_1 = \{\vec{v}_1, \vec{v}_2\} = \left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\}$$

and

$$B_2 = \left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\}$$

Find the matrix M_{B_2,B_1} of T with respect to the bases B_1 and B_2.

Answer

$$M_{B_2B_1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$$

This page titled 5.E: Exercises is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) .

CHAPTER OVERVIEW

6: Eigenvalues and Eigenvectors

Note 6.1

Solve the matrix equation $Ax = \lambda x$.

This chapter constitutes the core of any first course on linear algebra: eigenvalues and eigenvectors play a crucial role in most real-world applications of the subject.

Example 6.1

In a population of rabbits,

1. half of the newborn rabbits survive their first year;
2. of those, half survive their second year;
3. the maximum life span is three years;
4. rabbits produce 0, 6, 8 baby rabbits in their first, second, and third years, respectively.

What is the *asymptotic* behavior of this system? What will the rabbit population look like in 100 years?

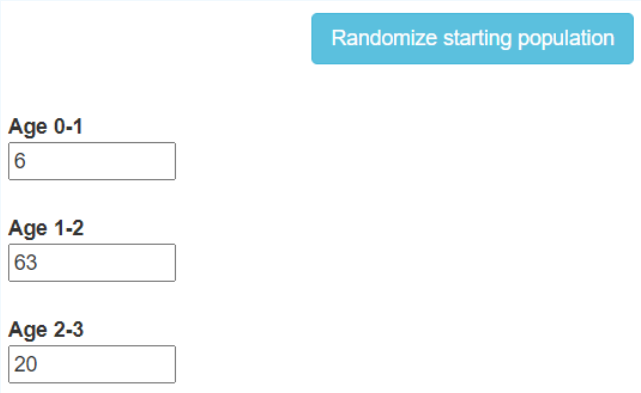


Figure 6.1: Left: the population of rabbits in a given year. Right: the proportions of rabbits in that year. Choose any values you like for the starting population, and click "Advance 1 year" several times. What do you notice about the long-term behavior of the ratios? This phenomenon turns out to be due to eigenvectors.

In Section 5.1, we will define eigenvalues and eigenvectors, and show how to compute the latter; in Section 5.2 we will learn to compute the former. In Section 5.3 we introduce the notion of *similar* matrices, and demonstrate that similar matrices do indeed behave similarly. In Section 5.4 we study matrices that are similar to diagonal matrices and in Section 5.5 we study matrices that are similar to rotation-scaling matrices, thus gaining a solid geometric understanding of large classes of matrices. Finally, we spend Section 5.6 presenting a common kind of application of eigenvalues and eigenvectors to real-world problems, including searching the Internet using Google's PageRank algorithm.

6.1: Eigenvalues and Eigenvectors
6.2: The Characteristic Polynomial
6.3: Diagonalization
6.4: Complex Eigenvalues
6.5: Stochastic Matrices
6.6: Similarity

This page titled 6: Eigenvalues and Eigenvectors is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

6.1: Eigenvalues and Eigenvectors

Learning Objectives

1. Learn the definition of eigenvector and eigenvalue.
2. Learn to find eigenvectors and eigenvalues geometrically.
3. Learn to decide if a number is an eigenvalue of a matrix, and if so, how to find an associated eigenvector.
4. *Recipe:* find a basis for the λ-eigenspace.
5. *Pictures:* whether or not a vector is an eigenvector, eigenvectors of standard matrix transformations.
6. *Theorem:* the expanded invertible matrix theorem.
7. *Vocabulary word:* ***eigenspace***.
8. *Essential vocabulary words:* ***eigenvector***, ***eigenvalue***.

In this section, we define eigenvalues and eigenvectors. These form the most important facet of the structure theory of square matrices. As such, eigenvalues and eigenvectors tend to play a key role in the real-life applications of linear algebra.

Eigenvalues and Eigenvectors

Here is the most important definition in this text.

Definition 6.1.1: Eigenvector and Eigenvalue

Let A be an $n \times n$ matrix.

1. An ***eigenvector*** of A is a *nonzero* vector v in $\mathbb{R}^n$ such that $Av = \lambda v,$ for some scalar λ.
2. An ***eigenvalue*** of A is a scalar λ such that the equation $Av = \lambda v$ has a *nontrivial* solution.

If $Av = \lambda v$ for $v \neq 0,$ we say that λ is the ***eigenvalue for*** v, and that v is an ***eigenvector for*** λ.

The German prefix "eigen" roughly translates to "self" or "own". An eigenvector of A is a vector that is taken to a multiple of itself by the matrix transformation $T(x) = Ax,$ which perhaps explains the terminology. On the other hand, "eigen" is often translated as "characteristic"; we may think of an eigenvector as describing an intrinsic, or characteristic, property of A.

Note 6.1.1

Eigenvalues and eigenvectors are only for square matrices.

Note 6.1.2

Eigenvectors are *by definition nonzero*. Eigenvalues may be equal to zero.

We do not consider the zero vector to be an eigenvector: since $A0 = 0 = \lambda 0$ for *every* scalar λ, the associated eigenvalue would be undefined.

If someone hands you a matrix A and a vector v, it is easy to check if v is an eigenvector of A: simply multiply v by A and see if Av is a scalar multiple of v. On the other hand, given just the matrix A, it is not obvious at all how to find the eigenvectors. We will learn how to do this in Section 5.2.

Example 6.1.1: Verifying eigenvectors

Consider the matrix

$$A = \begin{pmatrix} 2 & 2 \\ -4 & 8 \end{pmatrix} \qquad \text{and vectors} \qquad v = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad w = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Which are eigenvectors? What are their eigenvalues?

Solution

We have

$$Av = \begin{pmatrix} 2 & 2 \\ -4 & 8 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \end{pmatrix} = 4v.$$

Hence, v is an eigenvector of A, with eigenvalue $\lambda = 4$. On the other hand,

$$Aw = \begin{pmatrix} 2 & 2 \\ -4 & 8 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}.$$

which is not a scalar multiple of w. Hence, w is not an eigenvector of A.

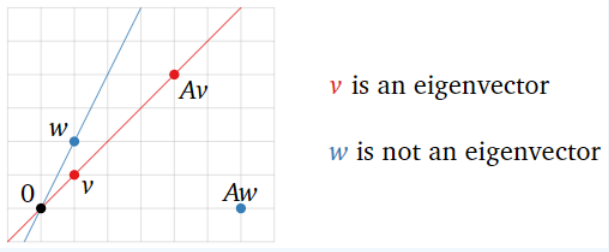


Figure 6.1.1

Example 6.1.2: Verifying eigenvectors

Consider the matrix

$$A = \begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix} \quad \text{and vectors} \quad v = \begin{pmatrix} 16 \\ 4 \\ 1 \end{pmatrix} \quad w = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}.$$

Which are eigenvectors? What are their eigenvalues?

Solution

We have

$$Av = \begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 16 \\ 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 32 \\ 8 \\ 2 \end{pmatrix} = 2v.$$

Hence, v is an eigenvector of A, with eigenvalue $\lambda = 2$. On the other hand,

$$Aw = \begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 28 \\ 1 \\ 1 \end{pmatrix},$$

which is not a scalar multiple of w. Hence, w is not an eigenvector of A.

Example 6.1.3: An eigenvector with eigenvalue 0

Let

$$A = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix} \qquad v = \begin{pmatrix} -3 \\ 1 \end{pmatrix}.$$

Is v an eigenvector of A? If so, what is its eigenvalue?

Solution

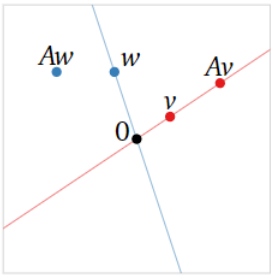
The product is

$$Av = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} -3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0v.$$

Hence, v is an eigenvector with eigenvalue *zero*.

As noted above, an eigen*value* is allowed to be zero, but an eigen*vector* is not.

To say that $Av = \lambda v$ means that Av and λv are *collinear with the origin*. So, an eigenvector of A is a nonzero vector v such that Av and v lie on the same line through the origin. In this case, Av is a scalar multiple of v; the eigenvalue is the scaling factor.



v is an eigenvector

w is not an eigenvector

Figure 6.1.2

For matrices that arise as the standard matrix of a linear transformation, it is often best to draw a picture, then find the eigenvectors and eigenvalues geometrically by studying which vectors are not moved off of their line. For a transformation that is defined geometrically, it is not necessary even to compute its matrix to find the eigenvectors and eigenvalues.

Example 6.1.4: Reflection

Here is an example of this. Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation that reflects over the line L defined by $y = -x$, and let A be the matrix for T. We will find the eigenvalues and eigenvectors of A without doing any computations.

This transformation is defined geometrically, so we draw a picture.

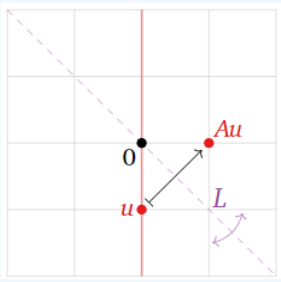


Figure 6.1.3

The vector u is not an eigenvector, because Au is not collinear with u and the origin.

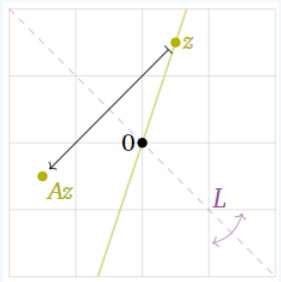


Figure 6.1.4

The vector z is not an eigenvector either.

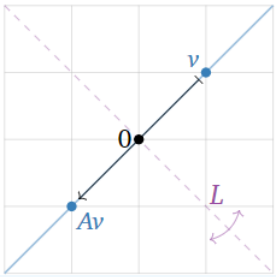


Figure 6.1.5

The vector v is an eigenvector because Av is collinear with v and the origin. The vector Av has the same length as v, but the opposite direction, so the associated eigenvalue is -1.

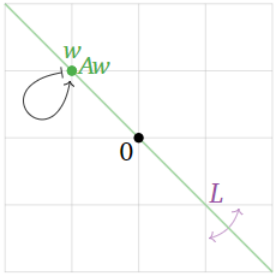


Figure 6.1.6

The vector w is an eigenvector because Aw is collinear with w and the origin: indeed, Aw is equal to w! This means that w is an eigenvector with eigenvalue 1.

It appears that all eigenvectors lie either on L, or on the line perpendicular to L. The vectors on L have eigenvalue 1, and the vectors perpendicular to L have eigenvalue -1.

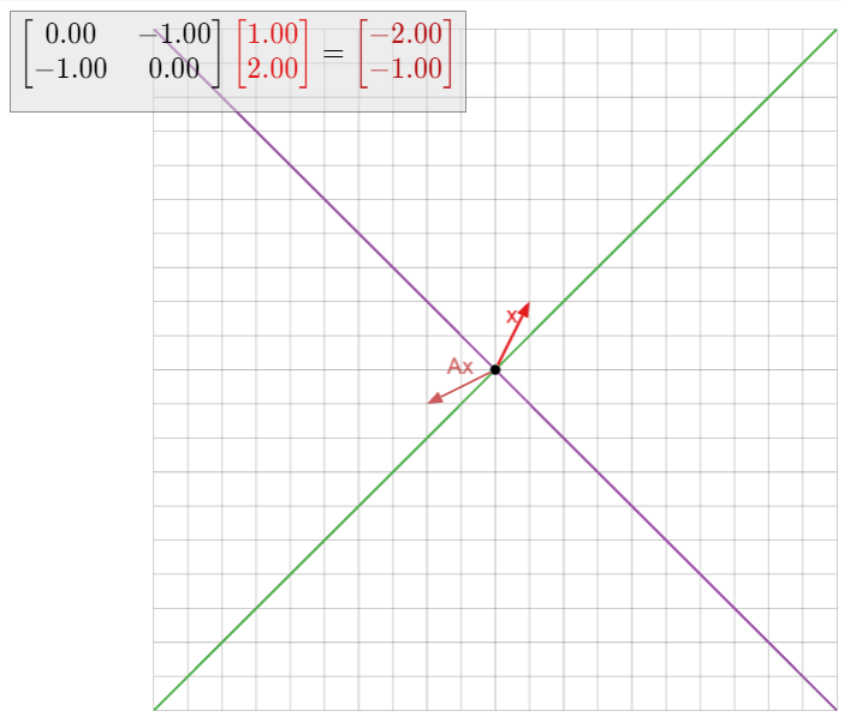


Figure 6.1.7: An eigenvector of A is a vector x such that Ax is collinear with x and the origin. Click and drag the head of x to convince yourself that all such vectors lie either on L, or on the line perpendicular to L.

We will now give five more examples of this nature

Example 6.1.5: Projection

Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation that projects a vector vertically onto the x-axis, and let A be the matrix for T. Find the eigenvalues and eigenvectors of A without doing any computations.

Solution

This transformation is defined geometrically, so we draw a picture.

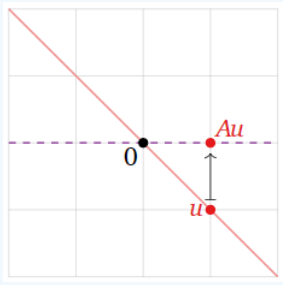


Figure 6.1.8

The vector u is not an eigenvector, because Au is not collinear with u and the origin.

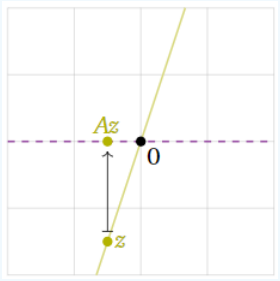


Figure 6.1.9

The vector z is not an eigenvector either.

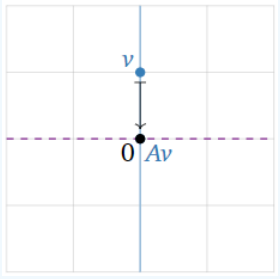


Figure 6.1.10

The vector v is an eigenvector. Indeed, Av is the zero vector, which is collinear with v and the origin; since $Av = 0v$, the associated eigenvalue is 0.

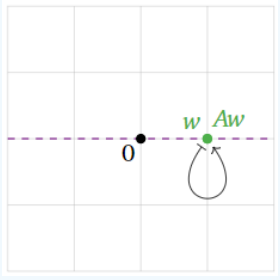


Figure 6.1.11

The vector w is an eigenvector because Aw is collinear with w and the origin: indeed, Aw is equal to w! This means that w is an eigenvector with eigenvalue 1.

It appears that all eigenvectors lie on the x-axis or the y-axis. The vectors on the x-axis have eigenvalue 1, and the vectors on the y-axis have eigenvalue 0.

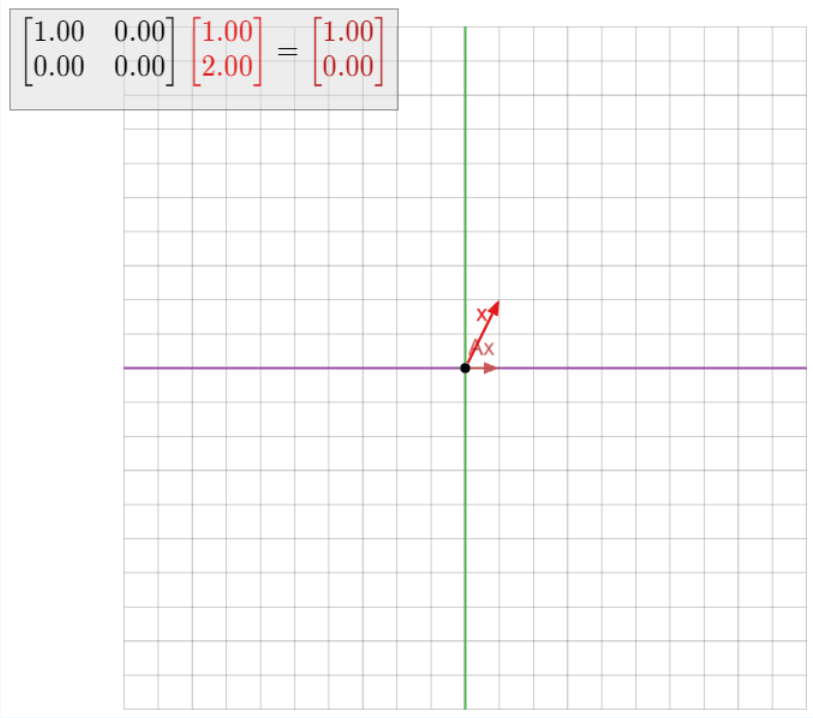


Figure 6.1.12: An eigenvector of A is a vector x such that Ax is collinear with x and the origin. Click and drag the head of x to convince yourself that all such vectors lie on the coordinate axes.

Example 6.1.6: Identity

Find all eigenvalues and eigenvectors of the identity matrix I_n.

Solution

The identity matrix has the property that $I_n v = v$ for all vectors v in $\mathbb{R}^n$. We can write this as $I_n v = 1 \cdot v$, so *every* nonzero vector is an eigenvector with eigenvalue 1.

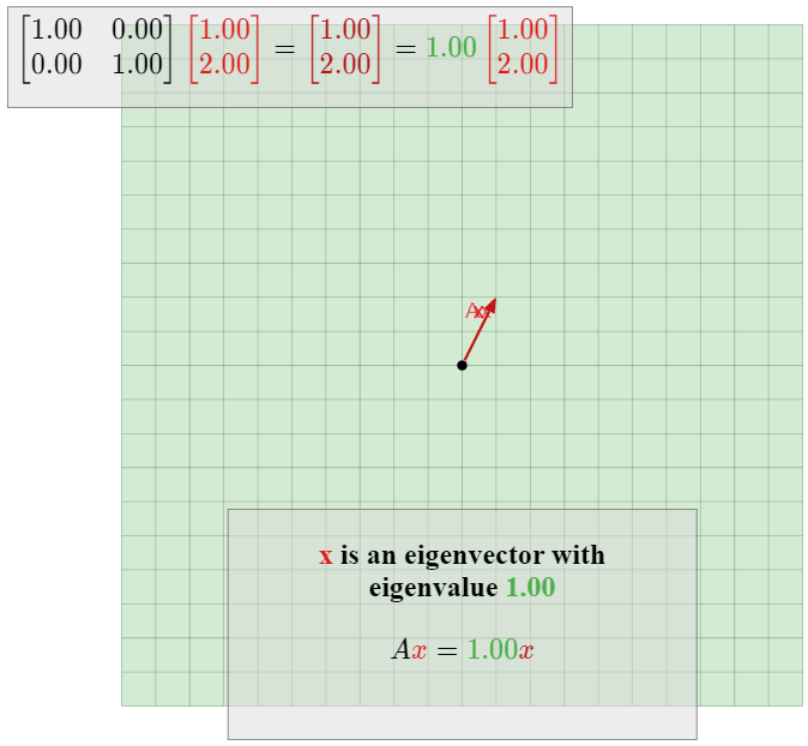


Figure 6.1.13: Every nonzero vector is an eigenvector of the identity matrix.

Example 6.1.7: Dilation

Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation that dilates by a factor of 1.5, and let A be the matrix for T. Find the eigenvalues and eigenvectors of A without doing any computations.

Solution

We have

$$Av = T(v) = 1.5v$$

for every vector v in $\mathbb{R}^2$. Therefore, by definition every nonzero vector is an eigenvector with eigenvalue 1.5.

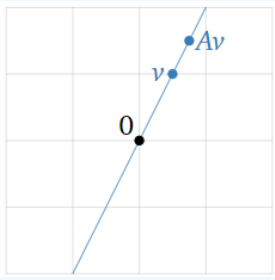


Figure 6.1.14

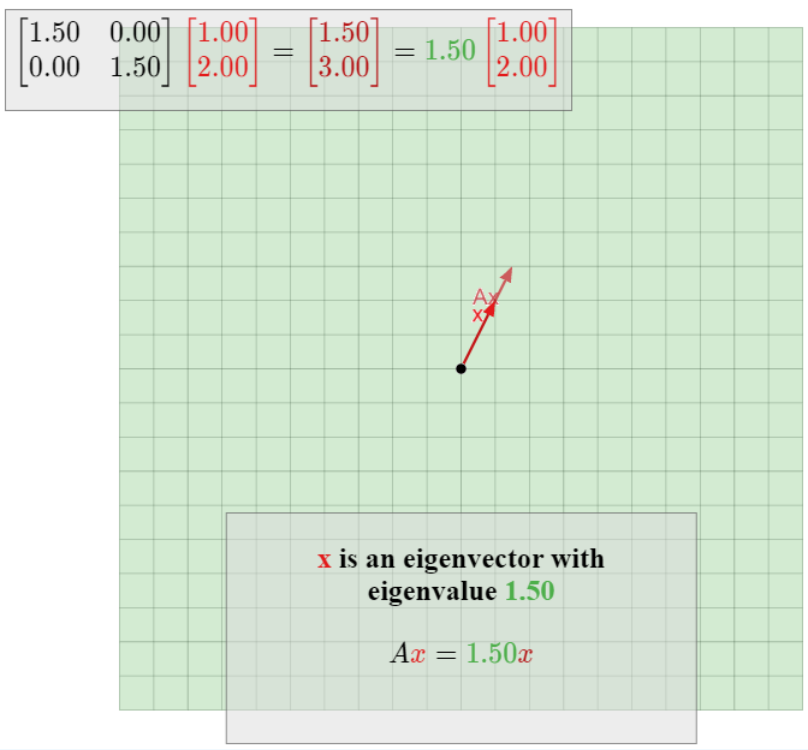


Figure 6.1.15: Every nonzero vector is an eigenvector of a dilation matrix.

Example 6.1.8: Shear

Let

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

and let $T(x) = Ax$, so T is a shear in the x-direction. Find the eigenvalues and eigenvectors of A without doing any computations.

Solution

In equations, we have

$$A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ y \end{pmatrix}.$$

This tells us that a shear takes a vector and adds its y-coordinate to its x-coordinate. Since the x-coordinate changes but not the y-coordinate, this tells us that any vector v with nonzero y-coordinate cannot be collinear with Av and the origin.

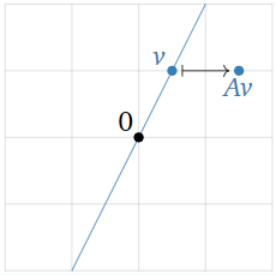


Figure 6.1.16

On the other hand, any vector v on the x-axis has zero y-coordinate, so it is not moved by A. Hence v is an eigenvector with eigenvalue 1.

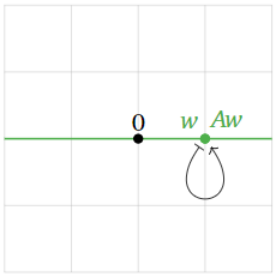


Figure 6.1.17

Accordingly, all eigenvectors of A lie on the x-axis, and have eigenvalue 1.



Figure 6.1.18: All eigenvectors of a shear lie on the x-axis. Click and drag the head of x to find the eigenvectors.

Example 6.1.9: Rotation

Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation that rotates counterclockwise by 90°, and let A be the matrix for T. Find the eigenvalues and eigenvectors of A without doing any computations.

Solution

If v is any nonzero vector, then Av is rotated by an angle of 90° from v. Therefore, Av is not on the same line as v, so v is not an eigenvector. And of course, the zero vector is never an eigenvector.

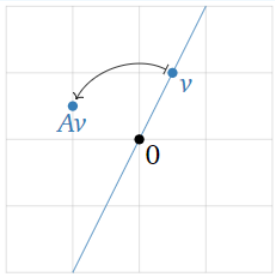


Figure 6.1.19

Therefore, this matrix *has no eigenvectors and eigenvalues.*

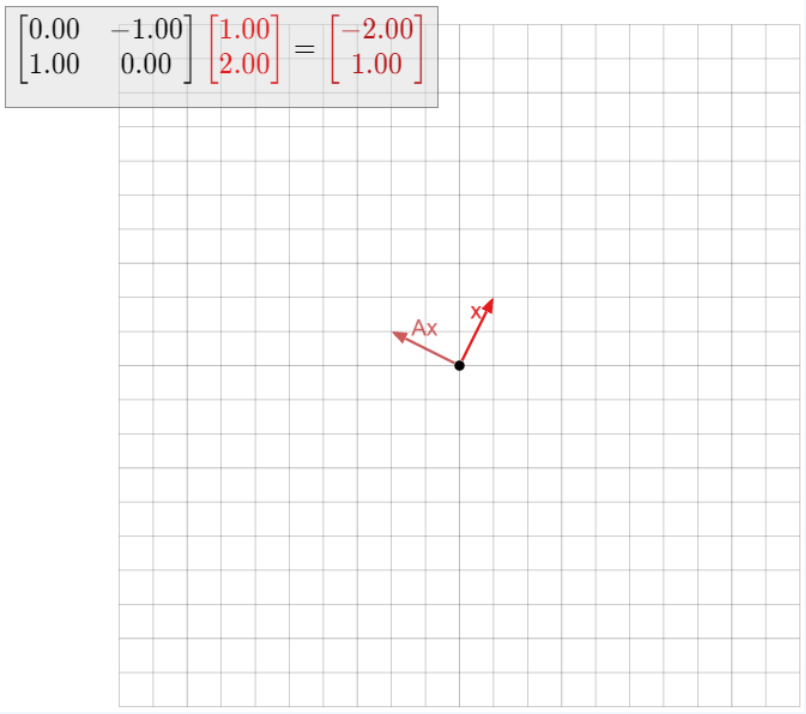


Figure 6.1.20: This rotation matrix has no eigenvectors. Click and drag the head of x to find one.

Here we mention one basic fact about eigenvectors.

Fact 6.1.1: Eigenvectors with Distinct Eigenvalues are Linearly Independent

Let $v_1, v_2, \ldots, v_k$ be eigenvectors of a matrix A, and suppose that the corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ are *distinct* (all different from each other). Then $\{v_1, v_2, \ldots, v_k\}$ is linearly independent.

Proof

Suppose that $\{v_1, v_2, \ldots, v_k\}$ were linearly *de*pendent. According to the increasing span criterion, Theorem 2.5.2 in Section 2.5, this means that for some j, the vector v_j is in $\text{Span}\{v_1, v_2, \ldots, v_{j-1}\}$. If we choose the *first* such j, then $\{v_1, v_2, \ldots, v_{j-1}\}$ is linearly independent. Note that $j > 1$ since $v_1 \neq 0$.

Since v_j is in $\text{Span}\{v_1, v_2, \ldots, v_{j-1}\}$, ,we can write

$$v_j = c_1 v_1 + c_2 v_2 + \cdots + c_{j-1} v_{j-1}$$

for some scalars $c_1, c_2, \ldots, c_{j-1}$. Multiplying both sides of the above equation by A gives

$$\begin{aligned} \lambda_j v_j = A v_j &= A\big(c_1 v_1 + c_2 v_2 + \cdots + c_{j-1} v_{j-1}\big) \\ &= c_1 A v_1 + c_2 A v_2 + \cdots + c_{j-1} A v_{j-1} \\ &= c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \cdots + c_{j-1} \lambda_{j-1} v_{j-1}. \end{aligned}$$

Subtracting λ_j times the first equation from the second gives

$$0 = \lambda_j v_j - \lambda_j v_j = c_1(\lambda_1 - \lambda_j) v_1 + c_2(\lambda_2 - \lambda_j) v_2 + \cdots + c_{j-1}(\lambda_{j-1} - \lambda_j) v_{j-1}.$$

Since $\lambda_i \neq \lambda_j$ for $i < j$, this is an equation of linear dependence among $v_1, v_2, \ldots, v_{j-1}$, which is impossible because those vectors are linearly independent. Therefore, $\{v_1, v_2, \ldots, v_k\}$ must have been linearly independent after all.

When $k = 2$, this says that if v_1, v_2 are eigenvectors with eigenvalues $\lambda_1 \neq \lambda_2$, then v_2 is not a multiple of v_1. In fact, any nonzero multiple cv_1 of v_1 is also an eigenvector with eigenvalue λ_1:

$$A(cv_1) = cAv_1 = c(\lambda_1 v_1) = \lambda_1(cv_1).$$

As a consequence of the above Fact 6.1.1, we have the following.

Note 6.1.3

An $n \times n$ matrix A has at most n eigenvalues.

Eigenspaces

Suppose that A is a square matrix. We already know how to check if a given vector is an eigenvector of A and in that case to find the eigenvalue. Our next goal is to check if a given real number is an eigenvalue of A and in that case to find all of the corresponding eigenvectors. Again this will be straightforward, but more involved. The only missing piece, then, will be to find the eigenvalues of A; this is the main content of Section 5.2.

Let A be an $n \times n$ matrix, and let λ be a scalar. The eigenvectors with eigenvalue λ, if any, are the nonzero solutions of the equation $Av = \lambda v$. We can rewrite this equation as follows:

$$\begin{aligned} & Av = \lambda v \\ \iff \quad & Av - \lambda v = 0 \\ \iff \quad & Av - \lambda I_n v = 0 \\ \iff \quad & (A - \lambda I_n)v = 0. \end{aligned}$$

Therefore, the eigenvectors of A with eigenvalue λ, if any, are the nontrivial solutions of the matrix equation $(A - \lambda I_n)v = 0$, i.e., the nonzero vectors in $\text{Nul}(A - \lambda I_n)$. If this equation has no nontrivial solutions, then λ is not an eigenvector of A.

The above observation is important because it says that *finding the eigenvectors for a given eigenvalue means solving a homogeneous system of equations.* For instance, if

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

then an eigenvector with eigenvalue λ is a nontrivial solution of the matrix equation

$$\begin{pmatrix} 7 & 1 & 3 \\ -3 & 2 & -3 \\ -3 & -2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

This translates to the system of equations

$$\begin{cases} 7x + y + 3z = \lambda x \\ -3x + 2y - 3z = \lambda y \\ -3x - 2y - z = \lambda z \end{cases} \longrightarrow \begin{cases} (7-\lambda)x + y + 3z = 0 \\ -3x + (2-\lambda)y - 3z = 0 \\ -3x - 2y + (-1-\lambda)z = 0. \end{cases}$$

This is the same as the homogeneous matrix equation

$$\begin{pmatrix} 7-\lambda & 1 & 3 \\ -3 & 2-\lambda & -3 \\ -3 & -2 & -1-\lambda \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0,$$

i.e., $(A - \lambda I_3)v = 0$.

Definition 6.1.2: λ-eigenspace

Let A be an $n \times n$ matrix, and let λ be an eigenvalue of A. The ***λ-eigenspace*** of A is the solution set of $(A - \lambda I_n)v = 0$, i.e., the subspace $\text{Nul}(A - \lambda I_n)$.

The λ-eigenspace is a subspace because it is the null space of a matrix, namely, the matrix $A - \lambda I_n$. This subspace consists of the zero vector and all eigenvectors of A with eigenvalue λ.

Note 6.1.4

Since a nonzero subspace is infinite, *every eigenvalue has infinitely many eigenvectors*. (For example, multiplying an eigenvector by a nonzero scalar gives another eigenvector.) On the other hand, there can be at most n linearly independent eigenvectors of an $n \times n$ matrix, since $\mathbb{R}^n$ has dimension n.

Example 6.1.10: Computing eigenspaces

For each of the numbers $\lambda = -2, 1, 3$, decide if λ is an eigenvalue of the matrix

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

and if so, compute a basis for the λ-eigenspace.

Solution

The number 3 is an eigenvalue of A if and only if $\text{Nul}(A - 3I_2)$ is nonzero. Hence, we have to solve the matrix equation $(A - 3I_2)v = 0$. We have

$$A - 3I_2 = \begin{pmatrix} 2 & -4 \\ -1 & -1 \end{pmatrix} - 3\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & -4 \\ -1 & -4 \end{pmatrix}.$$

The reduced row echelon form of this matrix is

$$\begin{pmatrix} 1 & 4 \\ 0 & 0 \end{pmatrix} \xrightarrow{\text{parametric form}} \begin{cases} x &= -4y \\ y &= y \end{cases} \xrightarrow{\text{parametric vector form}} \begin{pmatrix} x \\ y \end{pmatrix} = y\begin{pmatrix} -4 \\ 1 \end{pmatrix}.$$

Since y is a free variable, the null space of $A - 3I_2$ is nonzero, so 3 is an eigenvector. A basis for the 3-eigenspace is $\left\{\begin{pmatrix} -4 \\ 1 \end{pmatrix}\right\}$.

Concretely, we have shown that the eigenvectors of A with eigenvalue 3 are exactly the nonzero multiples of $\begin{pmatrix} -4 \\ 1 \end{pmatrix}$. In particular, $\begin{pmatrix} -4 \\ 1 \end{pmatrix}$ is an eigenvector, which we can verify:

$$\begin{pmatrix} 2 & -4 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} -4 \\ 1 \end{pmatrix} = \begin{pmatrix} -12 \\ 3 \end{pmatrix} = 3\begin{pmatrix} -4 \\ 1 \end{pmatrix}.$$

The number 1 is an eigenvalue of A if and only if $\text{Nul}(A - I_2)$ is nonzero. Hence, we have to solve the matrix equation $(A - I_2)v = 0$. We have

$$A - I_2 = \begin{pmatrix} 2 & -4 \\ -1 & -1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -4 \\ -1 & -2 \end{pmatrix}.$$

This matrix has determinant -6, so it is invertible. By Theorem 3.6.1 in Section 3.6, we have $\text{Nul}(A - I_2) = \{0\}$, so 1 is not an eigenvalue.

The eigenvectors of A with eigenvalue -2, if any, are the nonzero solutions of the matrix equation $(A + 2I_2)v = 0$. We have

$$A + 2I_2 = \begin{pmatrix} 2 & -4 \\ -1 & -1 \end{pmatrix} + 2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & -4 \\ -1 & 1 \end{pmatrix}.$$

The reduced row echelon form of this matrix is

$$\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \xrightarrow{\text{parametric form}} \begin{cases} x &= y \\ y &= y \end{cases} \xrightarrow{\text{parametric vector form}} \begin{pmatrix} x \\ y \end{pmatrix} = y\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Hence there exist eigenvectors with eigenvalue -2, namely, any nonzero multiple of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. A basis for the -2-eigenspace is $\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$.

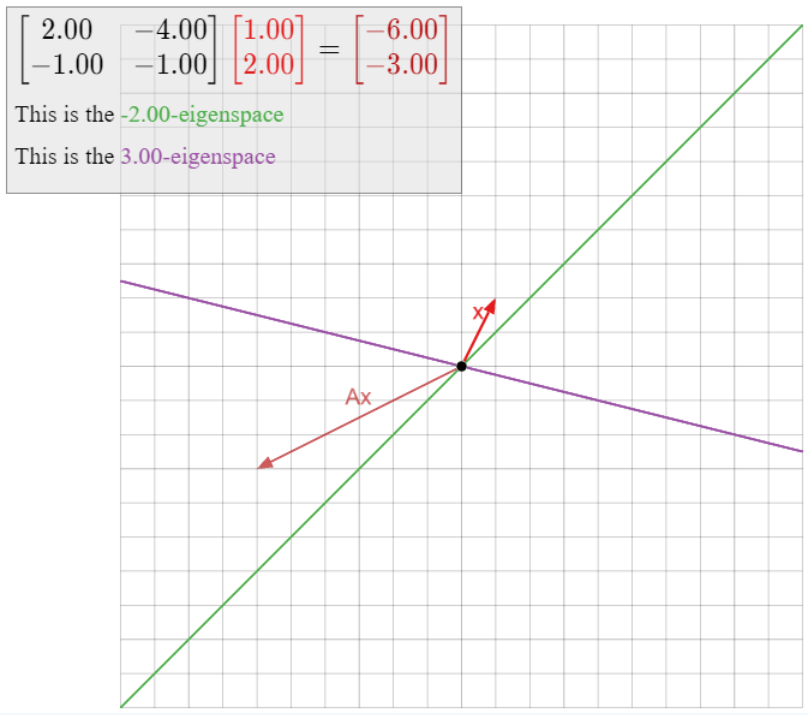


Figure 6.1.21: The 3-eigenspace is the line spanned by $\binom{-4}{1}$. This means that A scales every vector in that line by a factor of 3. Likewise, the -2-eigenspace is the line spanned by $\binom{1}{1}$. Click and drag the vector x around to see how A acts on that vector.

Example 6.1.11: Computing eigenspaces

For each of the numbers $\lambda = 0, \frac{1}{2}, 2$, decide if λ is an eigenvector of the matrix

$$A = \begin{pmatrix} 7/2 & 0 & 3 \\ -3/2 & 2 & -3 \\ -3/2 & 0 & -1 \end{pmatrix},$$

and if so, compute a basis for the λ-eigenspace.

Solution

The number 2 is an eigenvalue of A if and only if $\text{Nul}(A - 2I_3)$ is nonzero. Hence, we have to solve the matrix equation $(A - 2I_3)v = 0$. We have

$$A - 2I_3 = \begin{pmatrix} 7/2 & 0 & 3 \\ -3/2 & 2 & -3 \\ -3/2 & 0 & 1 \end{pmatrix} - 2\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3/2 & 0 & 3 \\ -3/2 & 0 & -3 \\ -3/2 & 0 & -3 \end{pmatrix}.$$

The reduced row echelon form of this matrix is

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{\text{parametric form}} \begin{cases} x &= -2z \\ y &= y \\ z &= z \end{cases} \xrightarrow{\text{parametric vector form}} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = y\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + z\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}.$$

The matrix $A - 2I_3$ has two free variables, so the null space of $A - 2I_3$ is nonzero, and thus 2 is an eigenvector. A basis for the 2-eigenspace is

$$\left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

This is a *plane* in $\mathbb{R}^3$.

The eigenvectors of A with eigenvalue $\frac{1}{2}$, if any, are the nonzero solutions of the matrix equation $(A-\frac{1}{2}I_3)v=0$. We have

$$A-\frac{1}{2}I_3=\begin{pmatrix}7/2&0&3\\-3/2&2&-3\\-3/2&0&1\end{pmatrix}-\frac{1}{2}\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}=\begin{pmatrix}3&0&3\\-3/2&3/2&-3\\-3/2&0&-3/2\end{pmatrix}.$$

The reduced row echelon form of this matrix is

$$\begin{pmatrix}1&0&1\\0&1&-1\\0&0&0\end{pmatrix}\xrightarrow{\text{parametric form}}\begin{cases}x&=&-z\\y&=&z\\z&=&z\end{cases}\xrightarrow{\text{parametric vector form}}\begin{pmatrix}x\\y\\z\end{pmatrix}=z\begin{pmatrix}-1\\1\\1\end{pmatrix}.$$

Hence there exist eigenvectors with eigenvalue $\frac{1}{2}$, so $\frac{1}{2}$ is an eigenvalue. A basis for the $\frac{1}{2}$-eigenspace is

$$\left\{\begin{pmatrix}-1\\1\\1\end{pmatrix}\right\}.$$

This is a *line* in $\mathbb{R}^3$.

The number 0 is an eigenvalue of A if and only if $\text{Nul}(A-0I_3)=\text{Nul}(A)$ is nonzero. This is the same as asking whether A is noninvertible, by Theorem 3.6.1 in Section 3.6. The determinant of A is $\det(A)=2\neq 0$, so A is invertible by the invertibility property, Proposition 4.1.2 in Section 4.1. It follows that 0 is not an eigenvalue of A.

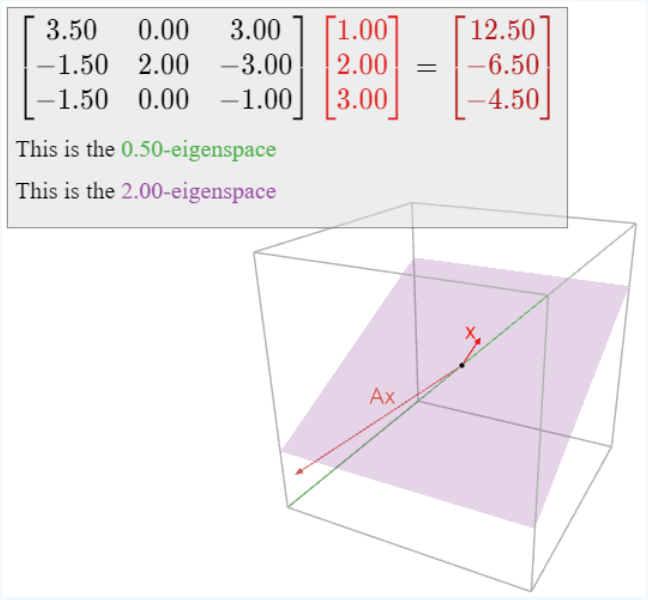


Figure 6.1.22: The 2-eigenspace is the violet plane. This means that A scales every vector in that plane by a factor of 2. The $\frac{1}{2}$-eigenspace is the green line. Click and drag the vector x around to see how A acts on that vector.

Example 6.1.12: Reflection

Let $T:\mathbb{R}^2\to\mathbb{R}^2$ be the linear transformation that reflects over the line L defined by $y=-x$, and let A be the matrix for T. Find all eigenspaces of A.

Solution

We showed in Example 6.1.4 that all eigenvectors with eigenvalue 1 lie on L, and all eigenvectors with eigenvalue -1 lie on the line $L^\perp$ that is perpendicular to L. Hence, L is the 1-eigenspace, and $L^\perp$ is the -1-eigenspace.

None of this required any computations, but we can verify our conclusions using algebra. First we compute the matrix A:

$$T\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}0\\-1\end{pmatrix}\quad T\begin{pmatrix}0\\1\end{pmatrix}=\begin{pmatrix}-1\\0\end{pmatrix}\quad\Longrightarrow\quad A=\begin{pmatrix}0&-1\\-1&0\end{pmatrix}.$$

Computing the 1-eigenspace means solving the matrix equation $(A-I_2)v=0$. We have

$$A-I_2=\begin{pmatrix}0&-1\\-1&0\end{pmatrix}-\begin{pmatrix}1&0\\0&1\end{pmatrix}=\begin{pmatrix}-1&-1\\-1&-1\end{pmatrix}\xrightarrow{\text{RREF}}\begin{pmatrix}1&1\\0&0\end{pmatrix}.$$

The parametric form of the solution set is $x=-y$, or equivalently, $y=-x$, which is exactly the equation for L. Computing the -1-eigenspace means solving the matrix equation $(A+I_2)v=0$; we have

$$A+I_2=\begin{pmatrix}0&-1\\-1&0\end{pmatrix}+\begin{pmatrix}1&0\\0&1\end{pmatrix}=\begin{pmatrix}1&-1\\-1&1\end{pmatrix}\xrightarrow{\text{RREF}}\begin{pmatrix}1&-1\\0&0\end{pmatrix}.$$

The parametric form of the solution set is $x=y$, or equivalently, $y=x$, which is exactly the equation for $L^\perp$.

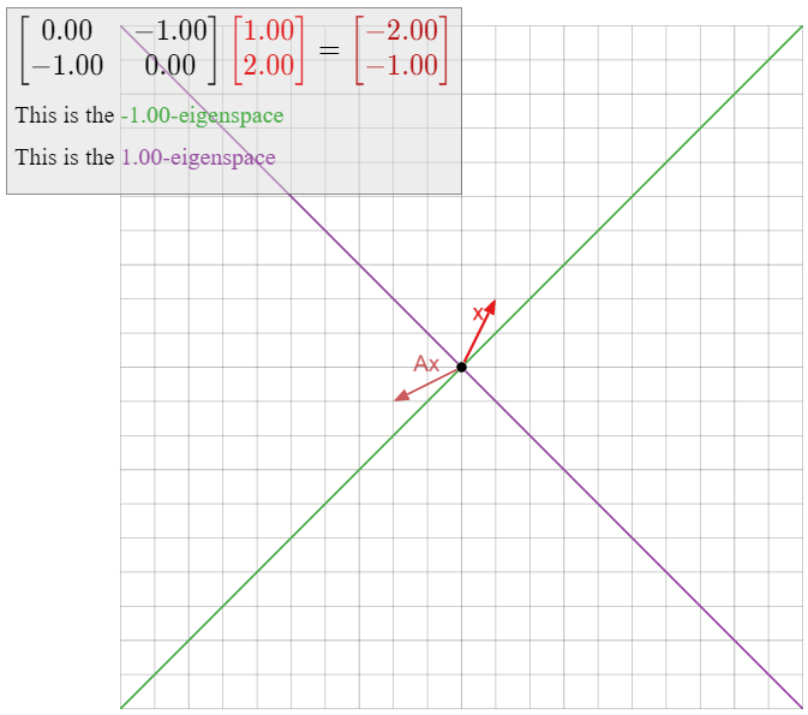


Figure 6.1.23: The violet line L is the 1-eigenspace, and the green line $L^\perp$ is the -1-eigenspace.

Recipe: Eigenspaces

Let A be an $n\times n$ matrix and let λ be a number.

1. λ is an eigenvalue of A if and only if $(A-\lambda I_n)v=0$ has a nontrivial solution, if and only if $\text{Nul}(A-\lambda I_n)\neq\{0\}$.
2. In this case, finding a basis for the λ-eigenspace of A means finding a basis for $\text{Nul}(A-\lambda I_n)$, which can be done by finding the parametric vector form of the solutions of the homogeneous system of equations $(A-\lambda I_n)v=0$.
3. The dimension of the λ-eigenspace of A is equal to the number of free variables in the system of equations $(A-\lambda I_n)v=0$, which is the number of columns of $A-\lambda I_n$ without pivots.
4. The eigenvectors with eigenvalue λ are the nonzero vectors in $\text{Nul}(A-\lambda I_n)$, or equivalently, the nontrivial solutions of $(A-\lambda I_n)v=0$.

We conclude with an observation about the 0-eigenspace of a matrix.

Fact 6.1.2

Let A be an $n \times n$ matrix.

1. The number 0 is an eigenvalue of A if and only if A is not invertible.
2. In this case, the 0-eigenspace of A is $\text{Nul}(A)$.

Proof

We know that 0 is an eigenvalue of A if and only if $\text{Nul}(A - 0I_n) = \text{Nul}(A)$ is nonzero, which is equivalent to the noninvertibility of A by Theorem 3.6.1 in Section 3.6. In this case, the 0-eigenspace is by definition $\text{Nul}(A - 0I_n) = \text{Nul}(A)$.

Concretely, an eigenvector with eigenvalue 0 is a nonzero vector v such that $Av = 0v$, i.e., such that $Av = 0$. These are exactly the nonzero vectors in the null space of A.

The Invertible Matrix Theorem: Addenda

We now have two new ways of saying that a matrix is invertible, so we add them to the invertible matrix theorem, Theorem 3.6.1 in Section 3.6.

Theorem 6.1.1: Invertible Matrix Theorem

Let A be an $n \times n$ matrix, and let $T: \mathbb{R}^n \to \mathbb{R}^n$ be the matrix transformation $T(x) = Ax$. The following statements are equivalent:

1. A is invertible.
2. A has n pivots.
3. $\text{Nul}(A) = \{0\}$.
4. The columns of A are linearly independent.
5. The columns of A span $\mathbb{R}^n$.
6. $Ax = b$ has a unique solution for each b in $\mathbb{R}^n$.
7. T is invertible.
8. T is one-to-one.
9. T is onto.
10. $\det(A) \neq 0$.
11. 0 is not an eigenvalue of A.

This page titled 6.1: Eigenvalues and Eigenvectors is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **5.1: Eigenvalues and Eigenvectors** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

6.2: The Characteristic Polynomial

Objectives

1. Learn that the eigenvalues of a triangular matrix are the diagonal entries.
2. Find all eigenvalues of a matrix using the characteristic polynomial.
3. Learn some strategies for finding the zeros of a polynomial.
4. *Recipe:* the characteristic polynomial of a 2×2 matrix.
5. *Vocabulary words:* ***characteristic polynomial***, ***trace***.

In Section 5.1 we discussed how to decide whether a given number λ is an eigenvalue of a matrix, and if so, how to find all of the associated eigenvectors. In this section, we will give a method for computing all of the *eigenvalues* of a matrix. This does not reduce to solving a system of linear equations: indeed, it requires solving a *nonlinear* equation in one variable, namely, finding the roots of the characteristic polynomial.

Definition 6.2.1: Characteristic Polynomial

Let A be an $n \times n$ matrix. The ***characteristic polynomial*** of A is the function $f(\lambda)$ given by

$$f(\lambda) = \det(A - \lambda I_n).$$

We will see below, Theorem 6.2.2, that the characteristic polynomial is in fact a polynomial. Finding the characterestic polynomial means computing the determinant of the matrix $A - \lambda I_n$, whose entries contain the unknown λ.

Example 6.2.1

Find the characteristic polynomial of the matrix

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}.$$

Solution

We have

$$\begin{aligned} f(\lambda) = \det(A - \lambda I_2) &= \det\left(\begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right) \\ &= \det\begin{pmatrix} 5-\lambda & 2 \\ 2 & 1-\lambda \end{pmatrix} \\ &= (5-\lambda)(1-\lambda) - 2 \cdot 2 = \lambda^2 - 6\lambda + 1. \end{aligned}$$

Example 6.2.2

Find the characteristic polynomial of the matrix

$$A = \begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

Solution

We compute the determinant by expanding cofactors along the third column:

$$\begin{aligned} f(\lambda) = \det(A - \lambda I_3) &= \det \begin{pmatrix} -\lambda & 6 & 8 \\ \frac{1}{2} & -\lambda & 0 \\ 0 & \frac{1}{2} & -\lambda \end{pmatrix} \\ &= 8\left(\frac{1}{4} - 0 \cdot -\lambda\right) - \lambda\left(\lambda^2 - 6 \cdot \frac{1}{2}\right) \\ &= -\lambda^3 + 3\lambda + 2. \end{aligned}$$

The point of the characteristic polynomial is that we can use it to compute eigenvalues.

Theorem 6.2.1: Eigenvalues are Roots of the Characteristic Polynomial

Let A be an $n \times n$ matrix, and let $f(\lambda) = \det(A - \lambda I_n)$ be its characteristic polynomial. Then a number λ_0 is an eigenvalue of A if and only if $f(\lambda_0) = 0$.

Proof

By the Theorem 5.1.1 in Section 5.1, the matrix equation $(A - \lambda_0 I_n)x = 0$ has a nontrivial solution if and only if $\det(A - \lambda_0 I_n) = 0$. Therefore,

$$\begin{aligned} \lambda_0 \text{ is an eigenvalue of } A &\iff Ax = \lambda_0 x \text{ has a nontrivial solution} \\ &\iff (A - \lambda_0 I_n)x = 0 \text{ has a nontrivial solution} \\ &\iff A - \lambda_0 I_n \text{ is not invertible} \\ &\iff \det(A - \lambda_0 I_n) = 0 \\ &\iff f(\lambda_0) = 0. \end{aligned}$$

Example 6.2.3: Finding eigenvalues

Find the eigenvalues and eigenvectors of the matrix

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}.$$

Solution

In the above Example 6.2.1 we computed the characteristic polynomial of A to be $f(\lambda) = \lambda^2 - 6\lambda + 1$. We can solve the equation $\lambda^2 - 6\lambda + 1 = 0$ using the quadratic formula:

$$\lambda = \frac{6 \pm \sqrt{36 - 4}}{2} = 3 \pm 2\sqrt{2}.$$

Therefore, the eigenvalues are $3 + 2\sqrt{2}$ and $3 - 2\sqrt{2}$.

To compute the eigenvectors, we solve the homogeneous system of equations $(A - \lambda I_2)x = 0$ for each eigenvalue λ. When $\lambda = 3 + 2\sqrt{2}$, we have

$$\begin{aligned} A - (3 + \sqrt{2})I_2 &= \begin{pmatrix} 2 - 2\sqrt{2} & 2 \\ 2 & -2 - 2\sqrt{2} \end{pmatrix} \\ &\xrightarrow{R_1 = R_1 \times (2 + 2\sqrt{2})} \begin{pmatrix} -4 & 4 + 4\sqrt{2} \\ 2 & -2 - 2\sqrt{2} \end{pmatrix} \\ &\xrightarrow{R_2 = R_2 + R_1/2} \begin{pmatrix} -4 & 4 + 4\sqrt{2} \\ 0 & 0 \end{pmatrix} \\ &\xrightarrow{R_1 = R_1 \div -4} \begin{pmatrix} 1 & -1 - \sqrt{2} \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

The parametric form of the general solution is $x = (1 + \sqrt{2})y$, so the $(3 + 2\sqrt{2})$-eigenspace is the line spanned by $\begin{pmatrix} 1 + \sqrt{2} \\ 1 \end{pmatrix}$. We compute in the same way that the $(3 - 2\sqrt{2})$-eigenspace is the line spanned by $\begin{pmatrix} 1 - \sqrt{2} \\ 1 \end{pmatrix}$.

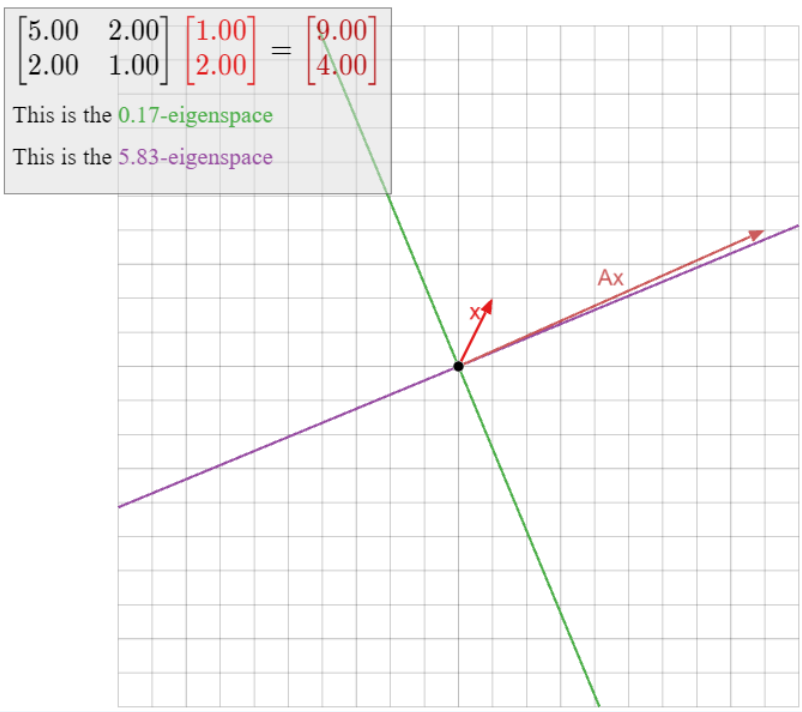


Figure 6.2.1: The green line is the $(3-2\sqrt{2})$-eigenspace, and the violet line is the $(3+2\sqrt{2})$-eigenspace.

Example 6.2.4: Finding eigenvalues

Find the eigenvalues and eigenvectors of the matrix

$$A=\begin{pmatrix}0 & 6 & 8\\ \frac{1}{2} & 0 & 0\\ 0 & \frac{1}{2} & 0\end{pmatrix}.$$

Solution

In the above Example 6.2.2 we computed the characteristic polynomial of A to be $f(\lambda)=-\lambda^3+3\lambda+2$. We eyeball that $f(2)=-8+3\cdot 2+2=0$. Thus $\lambda-2$ divides $f(\lambda)$; to find the other roots, we perform polynomial long division:

$$\frac{-\lambda^3+3\lambda+2}{\lambda-2}=-\lambda^2-2\lambda-1=-(\lambda+1)^2.$$

Therefore,

$$f(\lambda)=-(\lambda-2)(\lambda+1)^2,$$

so the only eigenvalues are $\lambda=2,-1$.

We compute the 2-eigenspace by solving the homogeneous system $(A-2I_3)x=0$. We have

$$A-2I_3=\begin{pmatrix}-2 & 6 & 8\\ \frac{1}{2} & -2 & 0\\ 0 & \frac{1}{2} & -2\end{pmatrix}\xrightarrow{\text{RREF}}\begin{pmatrix}1 & 0 & -16\\ 0 & 1 & -4\\ 0 & 0 & 0\end{pmatrix}.$$

The parametric form and parametric vector form of the solutions are:

$$\begin{cases}x &= 16z\\ y &= 4z\\ z &= z\end{cases}\longrightarrow\begin{pmatrix}x\\ y\\ z\end{pmatrix}=z\begin{pmatrix}16\\ 4\\ 1\end{pmatrix}.$$

Therefore, the 2-eigenspace is the line

$$\text{Span}\left\{\begin{pmatrix}16\\4\\1\end{pmatrix}\right\}.$$

We compute the -1-eigenspace by solving the homogeneous system $(A+I_3)x=0$. We have

$$A+I_3=\begin{pmatrix}1&6&8\\\frac{1}{2}&1&0\\0&\frac{1}{2}&1\end{pmatrix}\xrightarrow{\text{RREF}}\begin{pmatrix}1&0&-4\\0&1&2\\0&0&0\end{pmatrix}.$$

The parametric form and parametric vector form of the solutions are:

$$\begin{cases}z&=&4z\\y&=&-2z\\z&=&z\end{cases}\longrightarrow\begin{pmatrix}x\\y\\z\end{pmatrix}=z\begin{pmatrix}4\\-2\\1\end{pmatrix}.$$

Therefore, the -1-eigenspace is the line

$$\text{Span}\left\{\begin{pmatrix}4\\-2\\1\end{pmatrix}\right\}.$$

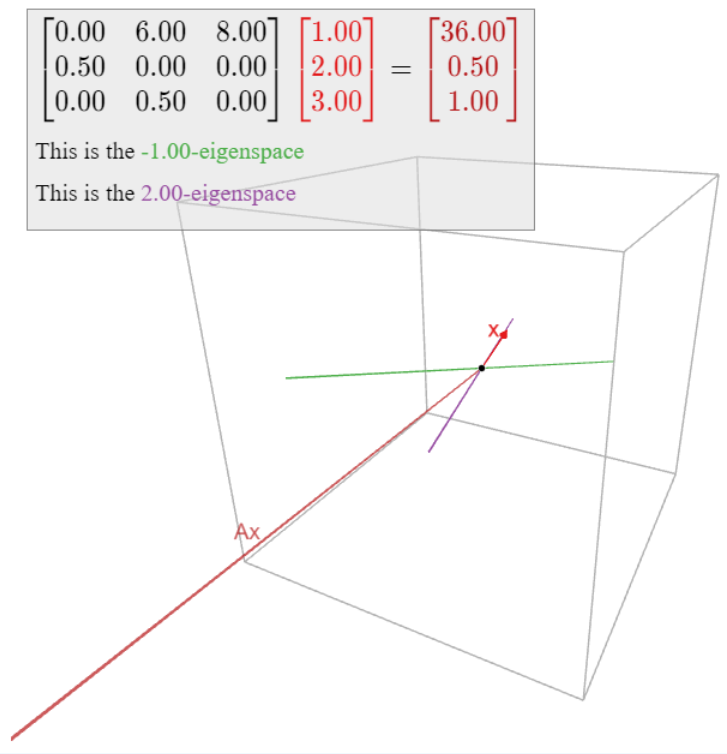


Figure 6.2.2: The green line is the -1-eigenspace, and the violet line is the 2-eigenspace.

Form of the Characteristic Polynomial

It is time that we justified the use of the term “polynomial.” First we need a vocabulary word.

Definition 6.2.2: Trace

The ***trace*** of a square matrix A is the number $\mathrm{Tr}(A)$ obtained by summing the diagonal entries of A:

$$\mathrm{Tr}\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n-1} & a_{n-1,n} \\ a_{n1} & a_{n2} & \cdots & a_{n,n-1} & a_{nn} \end{pmatrix} = a_{11}+a_{22}+\cdots+a_{nn}.$$

Theorem 6.2.2

Let A be an $n\times n$ matrix, and let $f(\lambda)=\det(A-\lambda I_n)$ be its characteristic polynomial. Then $f(\lambda)$ is a polynomial of degree n. Moreover, $f(\lambda)$ has the form

$$f(\lambda)=(-1)^n\lambda^n+(-1)^{n-1}\mathrm{Tr}(A)\lambda^{n-1}+\cdots+\det(A).$$

In other words, the coefficient of λ^{n-1} is $\pm\mathrm{Tr}(A)$, and the constant term is $\det(A)$ (the other coefficients are just numbers without names).

Proof

First we notice that

$$f(0)=\det(A-0I_n)=\det(A),$$

so that the constant term is always $\det(A)$.

We will prove the rest of the theorem only for 2×2 matrices; the reader is encouraged to complete the proof in general using cofactor expansions. We can write a 2×2 matrix as $A=\begin{pmatrix} a & b \\ c & d\end{pmatrix}$; then

$$\begin{aligned} f(\lambda) &= \det(A-\lambda I_2)=\det\begin{pmatrix} a-\lambda & b \\ c & d-\lambda\end{pmatrix}=(a-\lambda)(d-\lambda)-bc \\ &= \lambda^2-(a+d)\lambda+(ad-bc)=\lambda^2-\mathrm{Tr}(A)\lambda+\det(A). \end{aligned}$$

Recipe: The Characteristic Polynomial of a 2×2 Matrix

When $n=2$, the previous Theorem 6.2.2 tells us all of the coefficients of the characteristic polynomial:

$$f(\lambda)=\lambda^2-\mathrm{Tr}(A)\lambda+\det(A).$$

This is generally the fastest way to compute the characteristic polynomial of a 2×2 matrix.

Example 6.2.5

Find the characteristic polynomial of the matrix

$$A=\begin{pmatrix} 5 & 2 \\ 2 & 1\end{pmatrix}.$$

Solution

We have

$$f(\lambda)=\lambda^2-\mathrm{Tr}(A)\lambda+\det(A)=\lambda^2-(5+1)\lambda+(5\cdot 1-2\cdot 2)=\lambda^2-6\lambda+1,$$

as in the above Example 6.2.1.

Remark

By the above Theorem 6.2.2, the characteristic polynomial of an $n \times n$ matrix is a polynomial of degree n. Since a polynomial of degree n has at most n roots, this gives another proof of the fact that an $n \times n$ matrix has at most n eigenvalues. See Note 5.1.3 in Section 5.1.

Eigenvalues of a Triangular Matrix

It is easy to compute the determinant of an upper- or lower-triangular matrix; this makes it easy to find its eigenvalues as well.

Corollary 6.2.1

If A is an upper- or lower-triangular matrix, then the eigenvalues of A are its diagonal entries.

Proof

Suppose for simplicity that A is a 3×3 upper-triangular matrix:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix}.$$

Its characteristic polynomial is

$$f(\lambda) = \det(A - \lambda I_3) = \det \begin{pmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ 0 & a_{22} - \lambda & a_{23} \\ 0 & 0 & a_{33} - \lambda \end{pmatrix}.$$

This is also an upper-triangular matrix, so the determinant is the product of the diagonal entries:

$$f(\lambda) = (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda).$$

The zeros of this polynomial are exactly a_{11}, a_{22}, a_{33}.

Example 6.2.6

Find the eigenvalues of the matrix

$$A = \begin{pmatrix} 1 & 7 & 2 & 4 \\ 0 & 1 & 3 & 11 \\ 0 & 0 & \pi & 101 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Solution

The eigenvalues are the diagonal entries $1, \pi, 0$. (The eigenvalue 1 occurs twice, but it counts as one eigenvalue; in Section 5.4 we will define the notion of *algebraic multiplicity* of an eigenvalue.)

Factoring the Characteristic Polynomial

If A is an $n \times n$ matrix, then the characteristic polynomial $f(\lambda)$ has degree n by the above Theorem 6.2.2. When $n = 2$, one can use the quadratic formula to find the roots of $f(\lambda)$. There exist algebraic formulas for the roots of cubic and quartic polynomials, but these are generally too cumbersome to apply by hand. Even worse, it is known that there is no algebraic formula for the roots of a general polynomial of degree at least 5.

In practice, the roots of the characteristic polynomial are found numerically by computer. That said, there do exist methods for finding roots by hand. For instance, we have the following consequence of the rational root theorem (which we also call the rational root theorem):

Theorem 6.2.3: Rational Root Theorem

Suppose that A is an $n \times n$ matrix whose characteristic polynomial $f(\lambda)$ has integer (whole-number) entries. Then all rational roots of its characteristic polynomial are integer divisors of $\det(A)$.

For example, if A has integer entries, then its characteristic polynomial has integer coefficients. This gives us one way to find a root by hand, if A has an eigenvalue that is a rational number. Once we have found one root, then we can reduce the degree by polynomial long division.

Example 6.2.7

Find the eigenvalues of the matrix

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

Hint: one eigenvalue is an integer.

Solution

We compute the characteristic polynomial by expanding cofactors along the first row:

$$\begin{aligned} f(\lambda) = \det(A - \lambda I_3) &= \det \begin{pmatrix} 7-\lambda & 0 & 3 \\ -3 & 2-\lambda & -3 \\ -3 & 0 & -1-\lambda \end{pmatrix} \\ &= (7-\lambda)(2-\lambda)(-1-\lambda) + 3 \cdot 3(2-\lambda) \\ &= -\lambda^3 + 8\lambda^2 - 14\lambda + 4. \end{aligned}$$

The determinant of A is the constant term $f(0) = 4$; its integer divisors are $\pm 1, \pm 2, \pm 4$. We check which are roots:

$$f(1) = -3 \quad f(-1) = 27 \quad f(2) = 0 \quad f(-2) = 72 \quad f(4) = 12 \quad f(-4) = 252.$$

The only rational root of $f(\lambda)$ is $\lambda = 2$. We divide by $\lambda - 2$ using polynomial long division:

$$\frac{-\lambda^3 + 8\lambda^2 - 14\lambda + 4}{\lambda - 2} = -\lambda^2 + 6\lambda - 2.$$

We can use the quadratic formula to find the roots of the quotient:

$$\lambda = \frac{-6 \pm \sqrt{36 - 4 \cdot 2}}{-2} = 3 \pm \sqrt{7}.$$

We have factored f completely:

$$f(\lambda) = -(\lambda - 2)\big(\lambda - (3 + \sqrt{7})\big)\big(\lambda - (3 - \sqrt{7})\big).$$

Therefore, the eigenvalues of A are $2,\ 3 + \sqrt{7},\ 3 - \sqrt{7}$.

In the above example, we could have expanded cofactors along the second column to obtain

$$f(\lambda) = (2 - \lambda) \det \begin{pmatrix} 7-\lambda & 3 \\ -3 & -1-\lambda \end{pmatrix}.$$

Since $2 - \lambda$ was the only nonzero entry in its column, this expression already has the $2 - \lambda$ term factored out: the rational root theorem was not needed. The determinant in the above expression is the characteristic polynomial of the matrix $\begin{pmatrix} 7 & 3 \\ -3 & -1 \end{pmatrix}$, so we can compute it using the trace and determinant:

$$f(\lambda) = (2 - \lambda)\big(\lambda^2 - (7-1)\lambda + (-7+9)\big) = (2-\lambda)(\lambda^2 - 6\lambda + 2).$$

Example 6.2.8

Find the eigenvalues of the matrix

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

Solution

We compute the characteristic polynomial by expanding cofactors along the first row:

$$\begin{aligned} f(\lambda) = \det(A - \lambda I_3) &= \det \begin{pmatrix} 7-\lambda & 0 & 3 \\ -3 & 2-\lambda & -3 \\ 4 & 2 & -\lambda \end{pmatrix} \\ &= (7-\lambda)\big(-\lambda(2-\lambda)+6\big) + 3\big(-6-4(2-\lambda)\big) \\ &= -\lambda^3 + 9\lambda^2 - 8\lambda. \end{aligned}$$

The constant term is zero, so A has determinant zero. We factor out λ, then eyeball the roots of the quadratic factor:

$$f(\lambda) = -\lambda(\lambda^2 - 9\lambda + 8) = -\lambda(\lambda-1)(\lambda-8).$$

Therefore, the eigenvalues of A are $0, 1$, and 8.

Note 6.2.1: Finding Eigenvalues of a Matrix Larger than 2×2

Let A be an $n \times n$ matrix. Here are some strategies for factoring its characteristic polynomial $f(\lambda)$. First, you must find one eigenvalue:

1. Do not multiply out the characteristic polynomial if it is already partially factored! This happens if you expand cofactors along the second column in this Example 6.2.7.
2. If there is no constant term, you can factor out λ, as in this Example 6.2.8.
3. If the matrix is triangular, the roots are the diagonal entries.
4. Guess one eigenvalue using the Theorem 6.2.3: if $\det(A)$ is an integer, substitute all (positive and negative) divisors of $\det(A)$ into $f(\lambda)$.
5. Find an eigenvalue using the geometry of the matrix. For instance, a reflection, Example 5.1.4 in Section 5.1, has eigenvalues ± 1.

After obtaining an eigenvalue λ_1, use polynomial long division to compute $f(\lambda)/(\lambda - \lambda_1)$. This polynomial has lower degree. If $n = 3$ then this is a quadratic polynomial, to which you can apply the quadratic formula to find the remaining roots.

This page titled 6.2: The Characteristic Polynomial is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **5.2: The Characteristic Polynomial** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

6.3: Diagonalization

Objectives

1. Learn two main criteria for a matrix to be diagonalizable.
2. Develop a library of examples of matrices that are and are not diagonalizable.
3. *Recipes:* diagonalize a matrix, quickly compute powers of a matrix by diagonalization.
4. *Pictures:* the geometry of diagonal matrices, why a shear is not diagonalizable.
5. *Theorem:* the diagonalization theorem (two variants).
6. *Vocabulary words:* ***diagonalizable***, ***algebraic multiplicity***, ***geometric multiplicity***.

Diagonal matrices are the easiest kind of matrices to understand: they just scale the coordinate directions by their diagonal entries. In Section 5.3, we saw that similar matrices behave in the same way, with respect to different coordinate systems. Therefore, if a matrix is similar to a diagonal matrix, it is also relatively easy to understand. This section is devoted to the question: "When is a matrix similar to a diagonal matrix?" This section is devoted to the question: "When is a matrix *similar* to a diagonal matrix?" We will see that the algebra and geometry of such a matrix is relatively easy to understand.

Diagonalizability

Before answering the above question, first we give it a name.

Definition 6.3.1: Diagonalizable

An $n \times n$ matrix A is ***diagonalizable*** if it is similar to a diagonal matrix: that is, if there exists an invertible $n \times n$ matrix C and a diagonal matrix D such that

$$A = CDC^{-1}.$$

Example 6.3.1

Any diagonal matrix is D is diagonalizable because it is similar to itself, Proposition 5.3.1 in Section 5.3. For instance,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} = I_3 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} I_3^{-1}.$$

Example 6.3.2

Most of the examples in Section 5.3 involve diagonalizable matrices: The following are examples of diagonalizable matrices:

$$\begin{pmatrix} -12 & 15 \\ -10 & 13 \end{pmatrix} \quad \text{is diagonalizable because it equals} \quad \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix}^{-1}$$

$$\begin{pmatrix} 1/2 & 3/2 \\ 3/2 & 1/2 \end{pmatrix} \quad \text{is diagonalizable because it equals} \quad \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1}$$

$$\tfrac{1}{5}\begin{pmatrix} -8 & -9 \\ 6 & 13 \end{pmatrix} \quad \text{is diagonalizable because it equals} \quad \tfrac{1}{2}\begin{pmatrix} -1 & -3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \left(\tfrac{1}{2}\begin{pmatrix} -1 & -3 \\ 2 & 1 \end{pmatrix}\right)^{-1}$$

$$\begin{pmatrix} -1 & 0 & 0 \\ -1 & 0 & 2 \\ -1 & 1 & 1 \end{pmatrix} \quad \text{is diagonalizable because it equals} \quad \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix}^{-1}.$$

Example 6.3.3

If a matrix A is diagonalizable, and if B is similar to A, then B is diagonalizable as well by Proposition 5.3.1 in Section 5.3. as well. Indeed, if $A = CDC^{-1}$ for D diagonal, and $B = EAE^{-1}$, then

$$B = EAE^{-1} = E(CDC^{-1})E^{-1} = (EC)D(EC)^{-1},$$

so B is similar to D.

Powers of Diagonalizable Matrices

Multiplying diagonal matrices together just multiplies their diagonal entries:

$$\begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_2 & 0 \\ 0 & 0 & x_3 \end{pmatrix} \begin{pmatrix} y_1 & 0 & 0 \\ 0 & y_2 & 0 \\ 0 & 0 & y_3 \end{pmatrix} = \begin{pmatrix} x_1y_1 & 0 & 0 \\ 0 & x_2y_2 & 0 \\ 0 & 0 & x_3y_3 \end{pmatrix}.$$

Therefore, it is easy to take powers of a diagonal matrix:

$$\begin{pmatrix} x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & z \end{pmatrix}^n = \begin{pmatrix} x^n & 0 & 0 \\ 0 & y^n & 0 \\ 0 & 0 & z^n \end{pmatrix}.$$

By Fact 5.3.1 in Section 5.3, if $A = CDC^{-1}$ then $A^n = CD^nC^{-1}$, so it is also easy to take powers of *diagonalizable* matrices. This will be very important in applications to difference equations in Section 5.6. This is often very important in applications.

Recipe: Compute Powers of a Diagonalizable Matrix

If $A = CDC^{-1}$, where D is a diagonal matrix, then $A^n = CD^nC^{-1}$:

$$A = C \begin{pmatrix} x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & z \end{pmatrix} C^{-1} \quad \Longrightarrow \quad A^n = C \begin{pmatrix} x^n & 0 & 0 \\ 0 & y^n & 0 \\ 0 & 0 & z^n \end{pmatrix} C^{-1}.$$

Example 6.3.4

Let

$$A = \begin{pmatrix} 1/2 & 3/2 \\ 3/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1}.$$

Find a formula for A^n in which the entries are functions of n, where n is any positive whole number.

Solution

We have

$$\begin{aligned} A^n &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}^n \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2^n & 0 \\ 0 & (-1)^n \end{pmatrix} \frac{1}{-2} \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 2^n & (-1)^n \\ 2^n & (-1)^{n+1} \end{pmatrix} \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 2^n + (-1)^n & 2^n + (-1)^{n+1} \\ 2^n + (-1)^{n+1} & 2^n + (-1)^n \end{pmatrix}, \end{aligned}$$

where we used $(-1)^{n+2} = (-1)^2(-1)^n = (-1)^n$.

A fundamental question about a matrix is whether or not it is diagonalizable. The following is the primary criterion for diagonalizability. It shows that diagonalizability is an eigenvalue problem.

Theorem 6.3.1: Diagonalization Theorem

An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

In this case, $A = CDC^{-1}$ for

$$C = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \qquad D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where $v_1, v_2, \ldots, v_n$ are linearly independent eigenvectors, and $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the corresponding eigenvalues, in the same order.

Proof

First suppose that A has n linearly independent eigenvectors $v_1, v_2, \ldots, v_n$, with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Define C as above, so C is invertible by Theorem 5.1.1 in Section 5.1. Let $D = C^{-1}AC$, so $A = CDC^{-1}$. Multiplying by standard coordinate vectors, Fact 3.3.2 in Section 3.3, picks out the columns of C: we have $Ce_i = v_i$, so $e_i = C^{-1}v_i$. We multiply by the standard coordinate vectors to find the columns of D:

$$De_i = C^{-1}ACe_i = C^{-1}Av_i = C^{-1}\lambda_i v_i = \lambda_i C^{-1}v_i = \lambda_i e_i.$$

Therefore, the columns of D are multiples of the standard coordinate vectors:

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

Now suppose that $A = CDC^{-1}$, where C has columns $v_1, v_2, \ldots, v_n$, and D is diagonal with diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$. Since C is invertible, its columns are linearly independent. We have to show that v_i is an eigenvector of A with eigenvalue λ_i. We know that the standard coordinate vector e_i is an eigenvector of D with eigenvalue λ_i, so:

$$Av_i = CDC^{-1}v_i = CDe_i = C\lambda_i e_i = \lambda_i Ce_i = \lambda_i v_i.$$

By Fact 5.1.1 in Section 5.1, if an $n \times n$ matrix A has n *distinct* eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then a choice of corresponding eigenvectors $v_1, v_2, \ldots, v_n$ is automatically linearly independent.

Note 6.3.1

An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

Example 6.3.5: Easy Example

Apply Theorem 6.3.1 to the matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Solution

This diagonal matrix is in particular upper-triangular, so its eigenvalues are the diagonal entries $1, 2, 3$. The standard coordinate vectors are eigenvalues of a diagonal matrix:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = 2 \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 3 \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Therefore, the diagonalization theorem says that $A = CDC^{-1}$, where the columns of C are the standard coordinate vectors, and the D is the diagonal matrix with entries $1, 2, 3$:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}.$$

This just tells us that A is similar to itself.

Actually, the diagonalization theorem is not completely trivial even for diagonal matrices. If we put our eigenvalues in the order $3, 2, 1$, then the corresponding eigenvectors are e_3, e_2, e_1, so we also have that $A = C'D'(C')^{-1}$, where C' is the matrix with columns e_3, e_2, e_1, and D' is the diagonal matrix with entries $3, 2, 1$:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}^{-1}.$$

In particular, the matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

are similar to each other.

Note 6.3.2: Non-Uniqueness of Diagonalization

We saw in the above example that changing the order of the eigenvalues and eigenvectors produces a different diagonalization of the same matrix. There are generally many different ways to diagonalize a matrix, corresponding to different orderings of the eigenvalues of that matrix. The important thing is that the eigenvalues and eigenvectors have to be listed in the same order.

$$\begin{aligned} A &= \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix}^{-1} \\ &= \begin{pmatrix} | & | & | \\ v_3 & v_2 & v_1 \\ | & | & | \end{pmatrix} \begin{pmatrix} \lambda_3 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix} \begin{pmatrix} | & | & | \\ v_3 & v_2 & v_1 \\ | & | & | \end{pmatrix}^{-1}. \end{aligned}$$

There are other ways of finding different diagonalizations of the same matrix. For instance, you can scale one of the eigenvectors by a constant c:

$$A = \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} | & | & | \\ cv_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} | & | & | \\ cv_1 & v_2 & v_3 \\ | & | & | \end{pmatrix}^{-1}.$$

you can find a different basis entirely for an eigenspace of dimension at least 2, etc.

Example 6.3.6: A diagonalizable 2×2 matrix

Diagonalize the matrix

$$A = \begin{pmatrix} 1/2 & 3/2 \\ 3/2 & 1/2 \end{pmatrix}.$$

Solution

We need to find the eigenvalues and eigenvectors of A. First we compute the characteristic polynomial:

$$f(\lambda) = \lambda^2 - \operatorname{Tr}(A)\lambda + \det(A) = \lambda^2 - \lambda - 2 = (\lambda+1)(\lambda-2).$$

Therefore, the eigenvalues are -1 and 2. We need to compute eigenvectors for each eigenvalue. We start with $\lambda_1 = -1$:

$$(A + 1I_2)v = 0 \iff \begin{pmatrix} 3/2 & 3/2 \\ 3/2 & 3/2 \end{pmatrix} v = 0 \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} v = 0.$$

The parametric form is $x = -y$, so $v_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ is an eigenvector with eigenvalue λ_1. Now we find an eigenvector with eigenvalue $\lambda_2 = 2$:

$$(A - 2I_2)v = 0 \iff \begin{pmatrix} -3/2 & 3/2 \\ 3/2 & -3/2 \end{pmatrix} v = 0 \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} v = 0.$$

The parametric form is $x = y$, so $v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector with eigenvalue 2.

The eigenvectors v_1, v_2 are linearly independent, so Theorem 6.3.1 says that

$$A = CDC^{-1} \qquad \text{for} \qquad C = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Alternatively, if we choose 2 as our first eigenvalue, then

$$A = C'D'(C')^{-1} \qquad \text{for} \qquad C' = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad D' = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}.$$

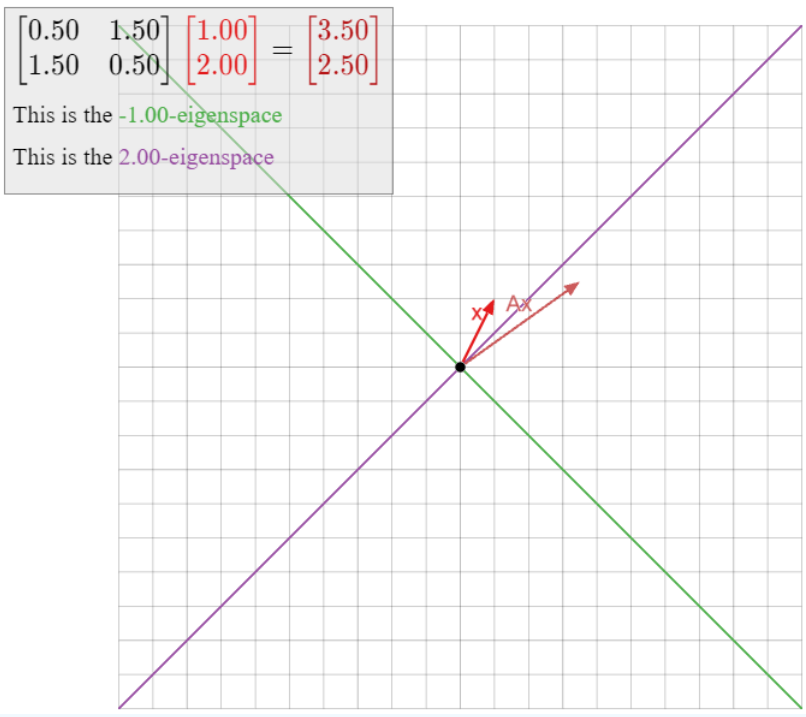


Figure 6.3.1: The green line is the -1-eigenspace of A, and the violet line is the 2-eigenspace. There are two linearly independent (noncollinear) eigenvectors visible in the picture: choose any nonzero vector on the green line, and any nonzero vector on the violet line.

Example 6.3.7: A diagonalizable 2×2 matrix with a zero eigenvector

Diagonalize the matrix

$$A = \begin{pmatrix} 2/3 & -4/3 \\ -2/3 & 4/3 \end{pmatrix}.$$

Solution

We need to find the eigenvalues and eigenvectors of A. First we compute the characteristic polynomial:

$$f(\lambda) = \lambda^2 - \operatorname{Tr}(A)\lambda + \det(A) = \lambda^2 - 2\lambda = \lambda(\lambda - 2).$$

Therefore, the eigenvalues are 0 and 2. We need to compute eigenvectors for each eigenvalue. We start with $\lambda_1 = 0$:

$$(A - 0I_2)v = 0 \iff \begin{pmatrix} 2/3 & -4/3 \\ -2/3 & 4/3 \end{pmatrix} v = 0 \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & -2 \\ 0 & 0 \end{pmatrix} v = 0.$$

The parametric form is $x = 2y$, so $v_1 = \binom{2}{1}$ is an eigenvector with eigenvalue λ_1. Now we find an eigenvector with eigenvalue $\lambda_2 = 2$:

$$(A - 2I_2)v = 0 \iff \begin{pmatrix} -4/3 & -4/3 \\ -2/3 & -2/3 \end{pmatrix} v = 0 \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} v = 0.$$

The parametric form is $x = -y$, so $v_2 = \binom{1}{-1}$ is an eigenvector with eigenvalue 2.

The eigenvectors v_1, v_2 are linearly independent, so the Theorem 6.3.1 says that

$$A = CDC^{-1} \qquad \text{for} \qquad C = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} \qquad D = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}.$$

Alternatively, if we choose 2 as our first eigenvalue, then

$$A = C'D'(C')^{-1} \qquad \text{for} \qquad C' = \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} \qquad D' = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

In the above example, the (non-invertible) matrix $A = \frac{1}{3}\begin{pmatrix} 2 & -4 \\ -2 & 4 \end{pmatrix}$ is similar to the diagonal matrix $D = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$. Since A is not invertible, zero is an eigenvalue by the invertible matrix theorem, Theorem 5.1.1 in Section 5.1, so one of the diagonal entries of D is necessarily zero. Also see Example 6.3.10 below.

Example 6.3.8: A diagonalizable 3×3 matrix

Diagonalize the matrix

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

Solution

We need to find the eigenvalues and eigenvectors of A. First we compute the characteristic polynomial by expanding cofactors along the third column:

$$\begin{aligned} f(\lambda) &= \det(A - \lambda I_3) = (1-\lambda)\det\left(\begin{pmatrix} 4 & -3 \\ 2 & -1 \end{pmatrix} - \lambda I_2\right) \\ &= (1-\lambda)(\lambda^2 - 3\lambda + 2) = -(\lambda-1)^2(\lambda-2). \end{aligned}$$

Therefore, the eigenvalues are 1 and 2. We need to compute eigenvectors for each eigenvalue. We start with $\lambda_1 = 1$:

$$(A - I_3)v = 0 \iff \begin{pmatrix} 3 & -3 & 0 \\ 2 & -2 & 0 \\ 1 & -1 & 0 \end{pmatrix} v = 0 \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} v = 0.$$

The parametric vector form is

$$\begin{cases} x &= y \\ y &= y \\ z &= z \end{cases} \implies \begin{pmatrix} x \\ y \\ z \end{pmatrix} + y\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + z\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Hence a basis for the 1-eigenspace is

$$\mathcal{B}_1 = \{v_1,\ v_2\} \quad \text{where} \quad v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Now we compute the eigenspace for $\lambda_2 = 2$:

$$(A - 2I_3)v = 0 \iff \begin{pmatrix} 2 & -3 & 0 \\ 2 & -3 & 0 \\ 1 & -1 & -1 \end{pmatrix} v = 0 \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix} v = 0$$

The parametric form is $x = 3z, y = 2z,$ so an eigenvector with eigenvalue 2 is

$$v_3 = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

The eigenvectors v_1, v_2, v_3 are linearly independent: v_1, v_2 form a basis for the 1-eigenspace, and v_3 is not contained in the 1-eigenspace because its eigenvalue is 2. Therefore, Theorem 6.3.1 says that

$$A = CDC^{-1} \qquad \text{for} \qquad C = \begin{pmatrix} 1 & 0 & 3 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

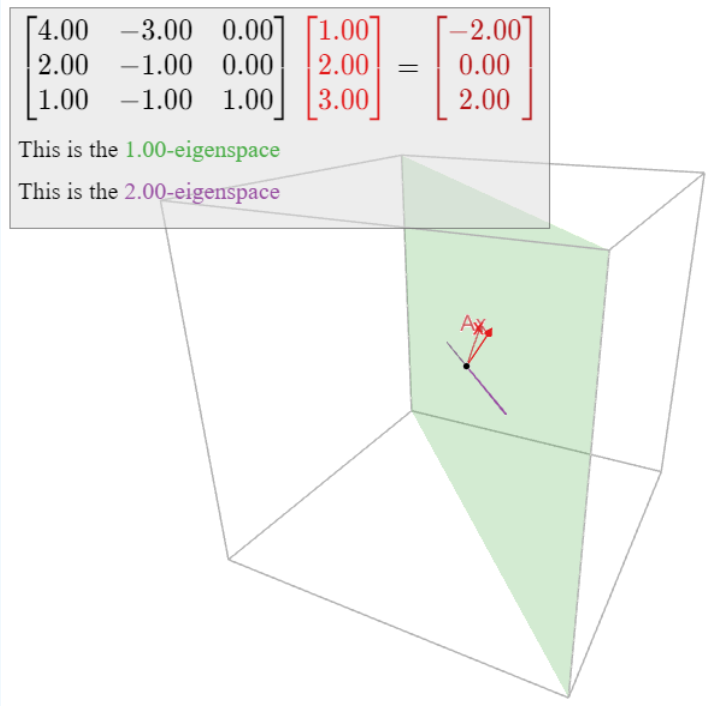


Figure 6.3.2: The green plane is the 1-eigenspace of A, and the violet line is the 2-eigenspace. There are three linearly independent eigenvectors visible in the picture: choose any two noncollinear vectors on the green plane, and any nonzero vector on the violet line.

Here is the procedure we used in the above examples.

Recipe: Diagonalization

Let A be an $n \times n$ matrix. To diagonalize A:

1. Find the eigenvalues of A using the characteristic polynomial.
2. For each eigenvalue λ of A, compute a basis $\mathcal{B}_\lambda$ for the λ-eigenspace.
3. If there are fewer than n total vectors in all of the eigenspace bases B_λ, then the matrix is not diagonalizable.
4. Otherwise, the n vectors $v_1, v_2, \ldots, v_n$ in the eigenspace bases are linearly independent, and $A = CDC^{-1}$ for

$$C = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where λ_i is the eigenvalue for v_i.

We will justify the linear independence assertion in part 4 in the proof of Theorem 6.3.3 below.

Example 6.3.9: A shear is not diagonalizable

Let

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

so $T(x) = Ax$ is a shear, Example 3.1.8 in Section 3.1. The characteristic polynomial of A is $f(\lambda) = (\lambda - 1)^2$, so the only eigenvalue of A is 1. We compute the 1-eigenspace:

$$(A - I_2)v = 0 \iff \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \iff y = 0.$$

In other words, the 1-eigenspace is exactly the x-axis, so *all* of the eigenvectors of A lie on the x-axis. It follows that A does *not* admit two linearly independent eigenvectors, so by Theorem 6.3.1, it is not diagonalizable.

In Example 5.1.8 in Section 5.1, we studied the eigenvalues of a shear geometrically; we reproduce the interactive demo here.



Figure 6.3.3: All eigenvectors of a shear lie on the x-axis.

Example 6.3.10: A projection is diagonalizable

Let L be a line through the origin in $\mathbb{R}^2$, and define $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ to be the transformation that sends a vector x to the closest point on L to x, as in the picture below.

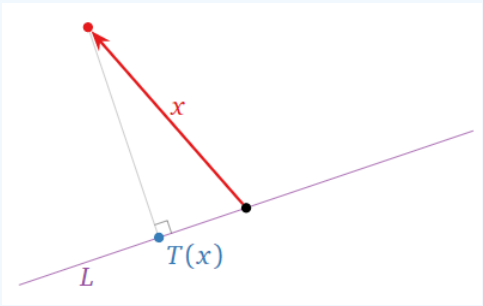


Figure 6.3.4

This is an example of an *orthogonal projection*. We will see in Section 6.3 that T is a linear transformation; let A be the matrix for T. Any vector on L is not moved by T because it is the closest point on L to itself: hence it is an eigenvector of A with eigenvalue 1. Let $L^\perp$ be the line perpendicular to L and passing through the origin. Any vector x on $L^\perp$ is closest to the zero vector on L, so a (nonzero) such vector is an eigenvector of A with eigenvalue 0. (See Example 5.1.5 in Section 5.1 for a

special case.) Since A has two distinct eigenvalues, it is diagonalizable; in fact, we know from Theorem 6.3.1 that A is similar to the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

Note that we never had to do any algebra! We know that A is diagonalizable for *geometric* reasons.

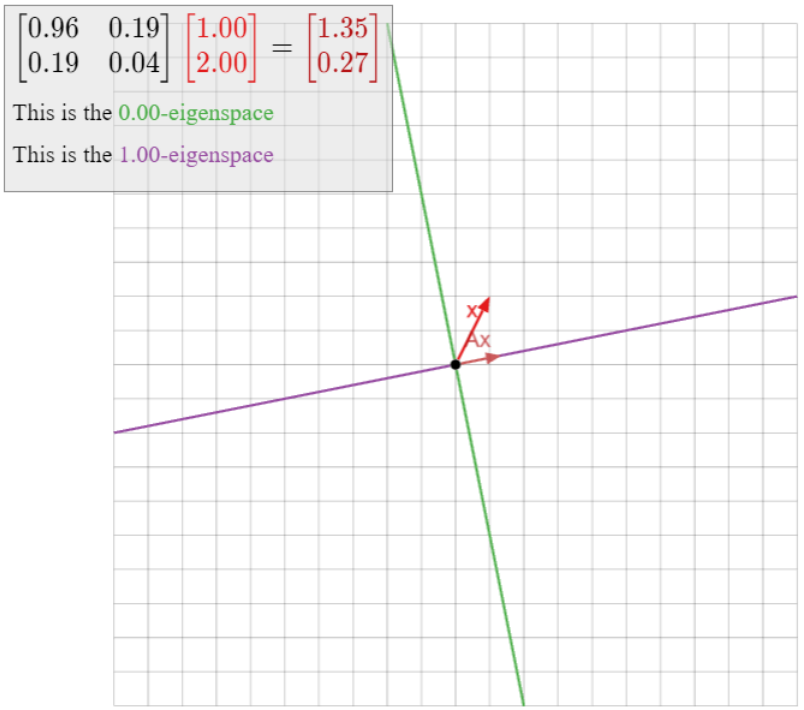


Figure 6.3.5: The line L (violet) is the 1-eigenspace of A, and $L^\perp$ (green) is the 0-eigenspace. Since there are linearly independent eigenvectors, we know that A is diagonalizable.

Example 6.3.11: A non-diagonalizable 3×3 matrix

Let

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

The characteristic polynomial of A is $f(\lambda) = -(\lambda-1)^2(\lambda-2)$, so the eigenvalues of A are 1 and 2. We compute the 1-eigenspace:

$$(A - I_3)v = 0 \iff \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0 \iff y = z = 0.$$

In other words, the 1-eigenspace is the x-axis. Similarly,

$$(A - 2I_3)v = 0 \iff \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0 \iff x = y = 0,$$

so the 2-eigenspace is the z-axis. In particular, all eigenvectors of A lie on the xz-plane, so there do not exist three linearly independent eigenvectors of A. By Theorem 6.3.1, the matrix A is not diagonalizable.

Notice that A contains a 2×2 block on its diagonal that looks like a shear:

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

This makes one suspect that such a matrix is not diagonalizable.

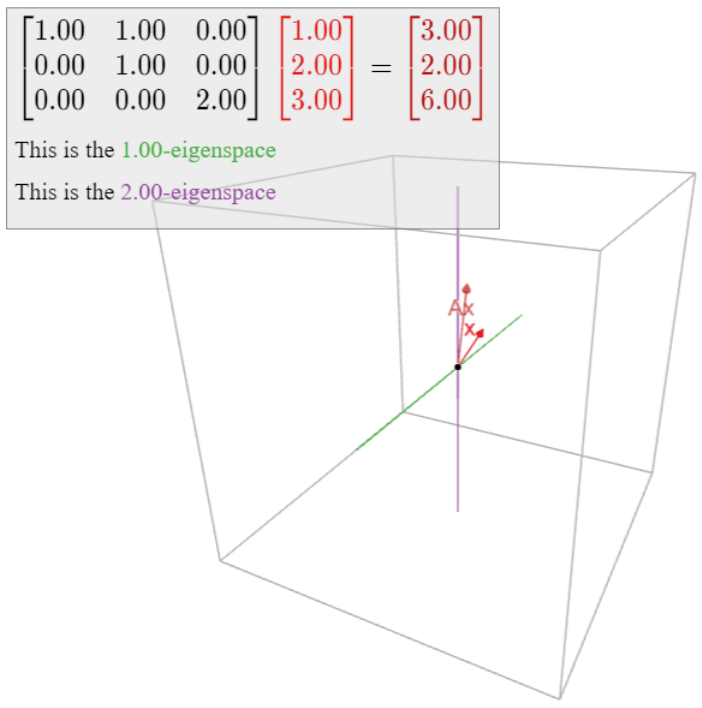


Figure 6.3.6: All eigenvectors of A lie on the x- and z-axes.

Example 6.3.12: A rotation matrix

Let

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

so $T(x) = Ax$ is the linear transformation that rotates counterclockwise by 90°. We saw in Example 5.1.9 in Section 5.1 that A does not have any eigenvectors at all. It follows that A is not diagonalizable.

$$\begin{bmatrix} 0.00 & -1.00 \\ 1.00 & 0.00 \end{bmatrix} \begin{bmatrix} 1.00 \\ 2.00 \end{bmatrix} = \begin{bmatrix} -2.00 \\ 1.00 \end{bmatrix}$$

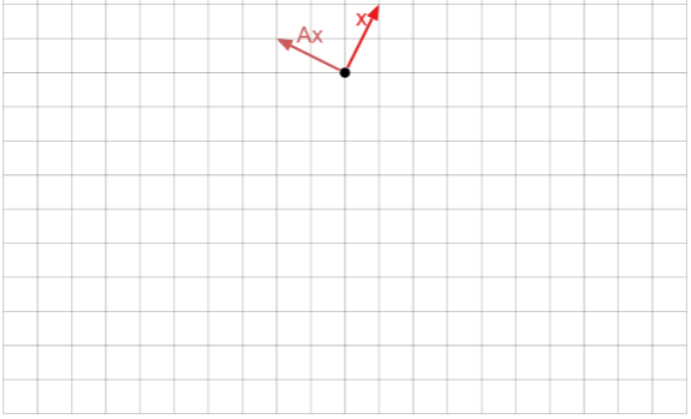


Figure 6.3.7: This rotation matrix has no eigenvectors.

The characteristic polynomial of A is $f(\lambda) = \lambda^2 + 1$, which of course does not have any real roots. If we allow *complex* numbers, however, then f has *two* roots, namely, $\pm i$, where $i = \sqrt{-1}$. Hence the matrix is diagonalizable if we allow ourselves to use complex numbers. We will treat this topic in detail in Section 5.5.

The following point is often a source of confusion.

Note 6.3.3: Diagonalizability has nothing to do with invertibility

Of the following matrices, the first is diagonalizable and invertible, the second is diagonalizable but not invertible, the third is invertible but not diagonalizable, and the fourth is neither invertible nor diagonalizable, as the reader can verify:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Remark: Non-diagonalizable 2×2 matrices with an eigenvalue

As in the above Example 6.3.9, one can check that the matrix

$$A_\lambda = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

is not diagonalizable for any number λ. We claim that any non-diagonalizable 2×2 matrix B with a real eigenvalue λ is similar to A_λ. Therefore, up to similarity, these are the only such examples.

To prove this, let B be such a matrix. Let v_1 be an eigenvector with eigenvalue λ, and let v_2 be any vector in $\mathbb{R}^2$ that is not collinear with v_1, so that $\{v_1, v_2\}$ forms a basis for $\mathbb{R}^2$. Let C be the matrix with columns v_1, v_2, and consider $A = C^{-1}BC$. We have $Ce_1 = v_1$ and $Ce_2 = v_2$, so $C^{-1}v_1 = e_1$ and $C^{-1}v_2 = e_2$. We can compute the first column of A as follows:

$$Ae_1 = C^{-1}BCe_1 = C^{-1}Bv_1 = C^{-1}\lambda v_1 = \lambda C^{-1}v_1 = \lambda e_1.$$

Therefore, A has the form

$$A = \begin{pmatrix} \lambda & b \\ 0 & d \end{pmatrix}.$$

Since A is similar to B, it also has only one eigenvalue λ; since A is upper-triangular, this implies $d = \lambda$, so

$$A = \begin{pmatrix} \lambda & b \\ 0 & \lambda \end{pmatrix}.$$

As B is not diagonalizable, we know A is not diagonal (B is similar to A), so $b \neq 0$. Now we observe that

$$\begin{pmatrix} 1/b & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda & b \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} 1/b & 0 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \lambda/b & 1 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} b & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} = A_\lambda.$$

We have shown that B is similar to A, which is similar to A_λ, so B is similar to A_λ by the transitivity property, Proposition 5.3.1 in Section 5.3, of similar matrices.

The Geometry of Diagonalizable Matrices

A diagonal matrix is easy to understand geometrically, as it just scales the coordinate axes:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = 2 \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 3 \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Therefore, we know from Section 5.3 that a diagonalizable matrix simply scales the "axes" with respect to a different coordinate system. Indeed, if $v_1, v_2, \ldots, v_n$ are linearly independent eigenvectors of an $n \times n$ matrix A, then A scales the v_i-direction by the eigenvalue λ_i.

In the following examples, we visualize the action of a diagonalizable matrix A in terms of its *dynamics*. In other words, we start with a collection of vectors (drawn as points), and we see where they move when we multiply them by A repeatedly.

Example 6.3.13: Eigenvalues $|\lambda_1| > 1$, $|\lambda_2| < 1$

Describe how the matrix

$$A = \frac{1}{10}\begin{pmatrix} 11 & 6 \\ 9 & 14 \end{pmatrix}$$

acts on the plane.

Solution

First we diagonalize A. The characteristic polynomial is

$$f(\lambda) = \lambda^2 - \operatorname{Tr}(A)\lambda + \det(A) = \lambda^2 - \frac{5}{2}\lambda + 1 = (\lambda - 2)\left(\lambda - \frac{1}{2}\right).$$

We compute the 2-eigenspace:

$$(A - 2I_3)v = 0 \iff \frac{1}{10}\begin{pmatrix} -9 & 6 \\ 9 & -6 \end{pmatrix} v = 0 \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & -2/3 \\ 0 & 0 \end{pmatrix} v = 0.$$

The parametric form of this equation is $x = 2/3y$, so one eigenvector is $v_1 = \begin{pmatrix} 2/3 \\ 1 \end{pmatrix}$. For the $1/2$-eigenspace, we have:

$$\left(A - \frac{1}{2}I_3\right)v = 0 \iff \frac{1}{10}\begin{pmatrix} 6 & 6 \\ 9 & 9 \end{pmatrix} v = 0 \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} v = 0.$$

The parametric form of this equation is $x = -y$, so an eigenvector is $v_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. It follows that $A = CDC^{-1}$, where

$$C = \begin{pmatrix} 2/3 & -1 \\ 1 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

The diagonal matrix D scales the x-coordinate by 2 and the y-coordinate by $1/2$. Therefore, it moves vectors closer to the x-axis and farther from the y-axis. In fact, since $(2x)(y/2) = xy$, multiplication by D does not move a point off of a hyperbola $xy = C$.

The matrix A does the same thing, in the v_1, v_2-coordinate system: multiplying a vector by A scales the v_1-coordinate by 2 and the v_2-coordinate by $1/2$. Therefore, A moves vectors closer to the 2-eigenspace and farther from the $1/2$-eigenspace.

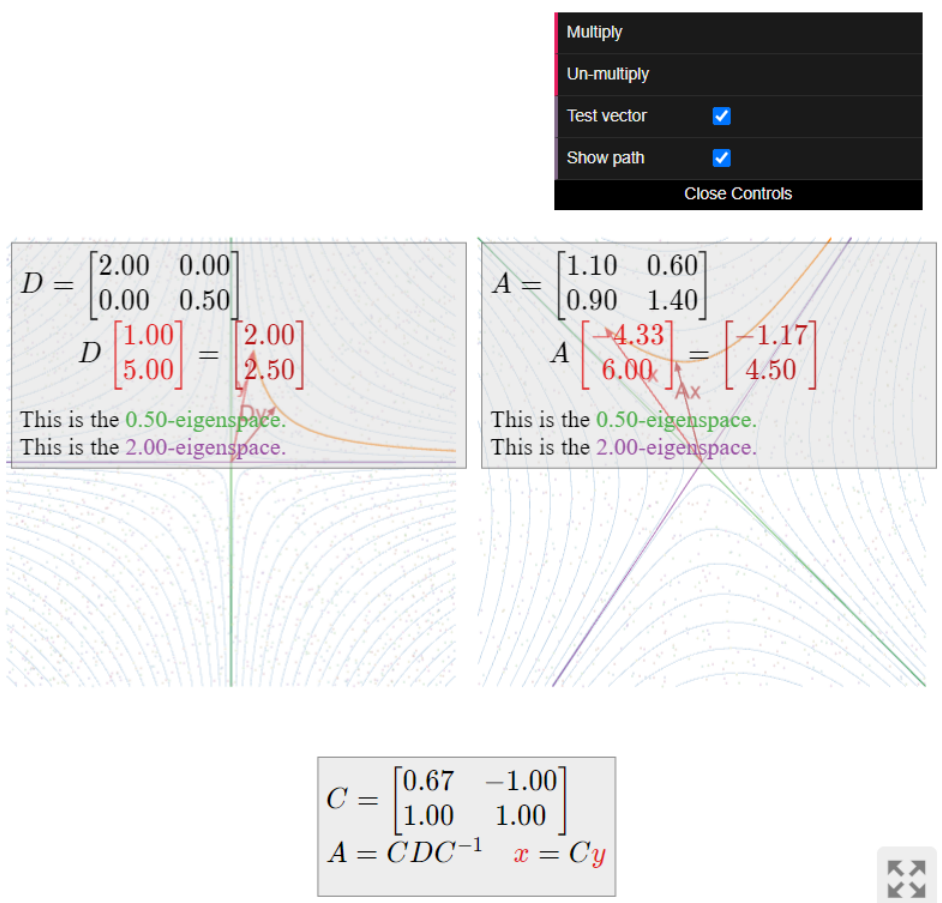


Figure 6.3.8: Dynamics of the matrices A and D. Click "multiply" to multiply the colored points by D on the left and A on the right.

Example 6.3.14: Eigenvalues $|\lambda_1| > 1$, $|\lambda_2| > 1$

Describe how the matrix

$$A = \frac{1}{5}\begin{pmatrix} 13 & -2 \\ -3 & 12 \end{pmatrix}$$

acts on the plane.

Solution

First we diagonalize A. The characteristic polynomial is

$$f(\lambda) = \lambda^2 - \mathrm{Tr}(A)\lambda + \det(A) = \lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3).$$

Next we compute the 2-eigenspace:

$$(A-2I_3)v=0 \iff \frac{1}{5}\begin{pmatrix}3&-2\\-3&2\end{pmatrix}v=0 \xrightarrow{\text{RREF}} \begin{pmatrix}1&-2/3\\0&0\end{pmatrix}v=0.$$

The parametric form of this equation is $x=2/3y$, so one eigenvector is $v_1=\binom{2/3}{1}$. For the 3-eigenspace, we have:

$$(A-3I_3)v=0 \iff \frac{1}{5}\begin{pmatrix}-2&-2\\-3&-3\end{pmatrix}v=0 \xrightarrow{\text{RREF}} \begin{pmatrix}1&1\\0&0\end{pmatrix}v=0.$$

The parametric form of this equation is $x=-y$, so an eigenvector is $v_2=\binom{-1}{1}$. It follows that $A=CDC^{-1}$, where

$$C=\begin{pmatrix}2/3&-1\\1&1\end{pmatrix} \qquad D=\begin{pmatrix}2&0\\0&3\end{pmatrix}.$$

The diagonal matrix D scales the x-coordinate by 2 and the y-coordinate by 3. Therefore, it moves vectors farther from both the x-axis and the y-axis, but faster in the y-direction than the x-direction.

The matrix A does the same thing, in the v_1, v_2-coordinate system: multiplying a vector by A scales the v_1-coordinate by 2 and the v_2-coordinate by 3. Therefore, A moves vectors farther from the 2-eigenspace and the 3-eigenspace, but faster in the v_2-direction than the v_1-direction.

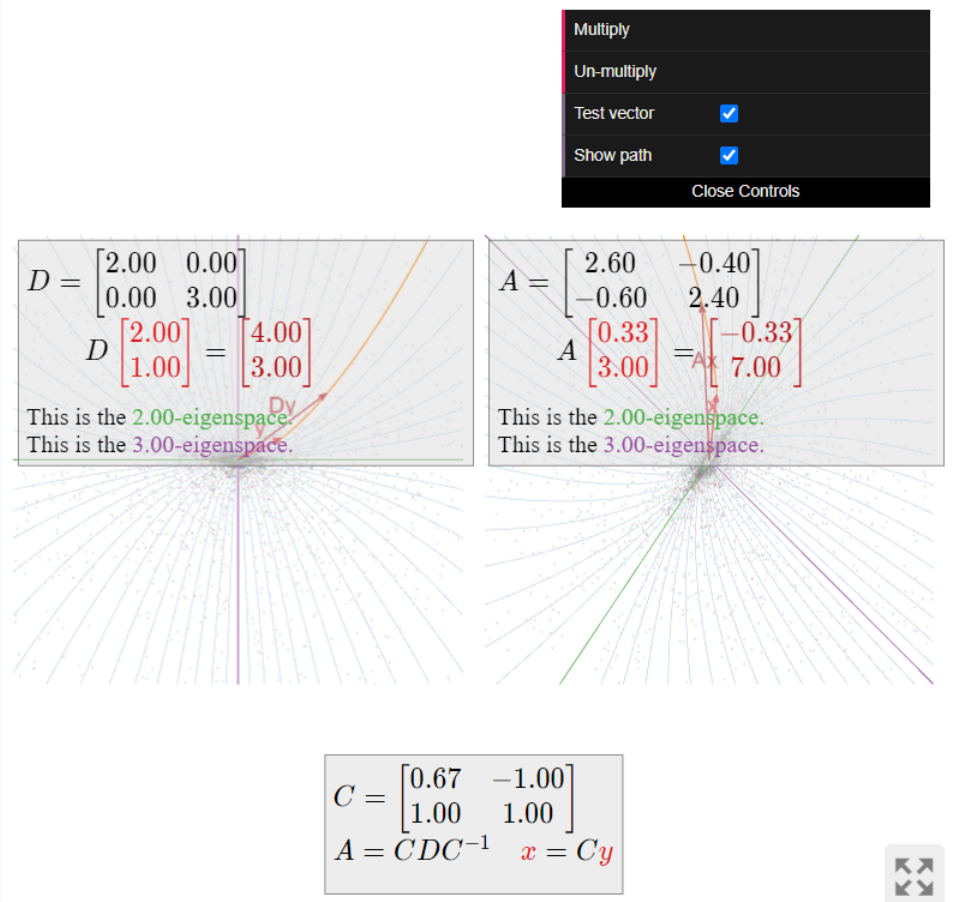


Figure 6.3.9: Dynamics of the matrices A and D. Click "multiply" to multiply the colored points by D on the left and A on the right.

Example 6.3.15: Eigenvalues $|\lambda_1|<1$, $|\lambda_2|<1$

Describe how the matrix

$$A'=\frac{1}{30}\begin{pmatrix}12&2\\3&13\end{pmatrix}$$

acts on the plane.

Solution

This is the inverse of the matrix A from the previous Example 6.3.14 In that example, we found $A = CDC^{-1}$ for

$$C = \begin{pmatrix} 2/3 & -1 \\ 1 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}.$$

Therefore, remembering that taking inverses reverses the order of multiplication, Fact 3.5.1 in Section 3.5, we have

$$A' = A^{-1} = (CDC^{-1})^{-1} = (C^{-1})^{-1} D^{-1} C^{-1} = C \begin{pmatrix} 1/2 & 0 \\ 0 & 1/3 \end{pmatrix} C^{-1}.$$

The diagonal matrix D^{-1} does the opposite of what D does: it scales the x-coordinate by $1/2$ and the y-coordinate by $1/3$. Therefore, it moves vectors closer to both coordinate axes, but faster in the y-direction. The matrix A' does the same thing, but with respect to the v_1, v_2-coordinate system.

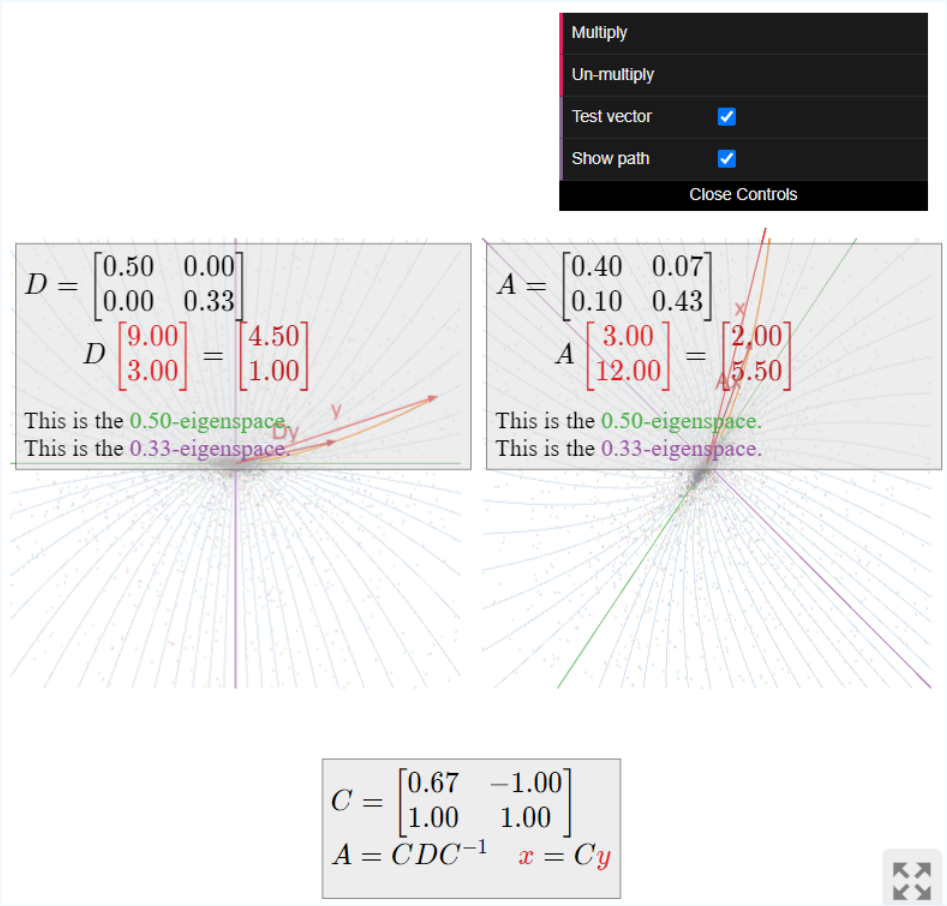


Figure 6.3.10: Dynamics of the matrices A' and D^{-1}. Click "multiply" to multiply the colored points by D^{-1} on the left and A' on the right.

Example 6.3.16: Eigenvalues $|\lambda_1| = 1$, $|\lambda_2| < 1$

Describe how the matrix

$$A = \frac{1}{6} \begin{pmatrix} 5 & -1 \\ -2 & 4 \end{pmatrix}$$

acts on the plane.

Solution

First we diagonalize A. The characteristic polynomial is

$$f(\lambda) = \lambda^2 - \mathrm{Tr}(A)\lambda + \det(A) = \lambda^2 - \frac{3}{2}\lambda + \frac{1}{2} = (\lambda - 1)\left(\lambda - \frac{1}{2}\right).$$

Next we compute the 1-eigenspace:

$$(A - I_3)v = 0 \iff \frac{1}{6}\begin{pmatrix} -1 & -1 \\ -2 & -2 \end{pmatrix} v = 0 \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} v = 0.$$

The parametric form of this equation is $x = -y$, so one eigenvector is $v_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. For the $1/2$-eigenspace, we have:

$$\left(A - \frac{1}{2}I_3\right)v = 0 \iff \frac{1}{6}\begin{pmatrix} 2 & -1 \\ -2 & 1 \end{pmatrix} v = 0 \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & -1/2 \\ 0 & 0 \end{pmatrix} v = 0.$$

The parametric form of this equation is $x = 1/2y$, so an eigenvector is $v_2 = \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}$. It follows that $A = CDC^{-1}$, where

$$C = \begin{pmatrix} -1 & 1/2 \\ 1 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

The diagonal matrix D scales the y-coordinate by $1/2$ and does not move the x-coordinate. Therefore, it simply moves vectors closer to the x-axis along vertical lines. The matrix A does the same thing, in the v_1, v_2-coordinate system: multiplying a vector by A scales the v_2-coordinate by $1/2$ and does not change the v_1-coordinate. Therefore, A "sucks vectors into the 1-eigenspace" along lines parallel to v_2.

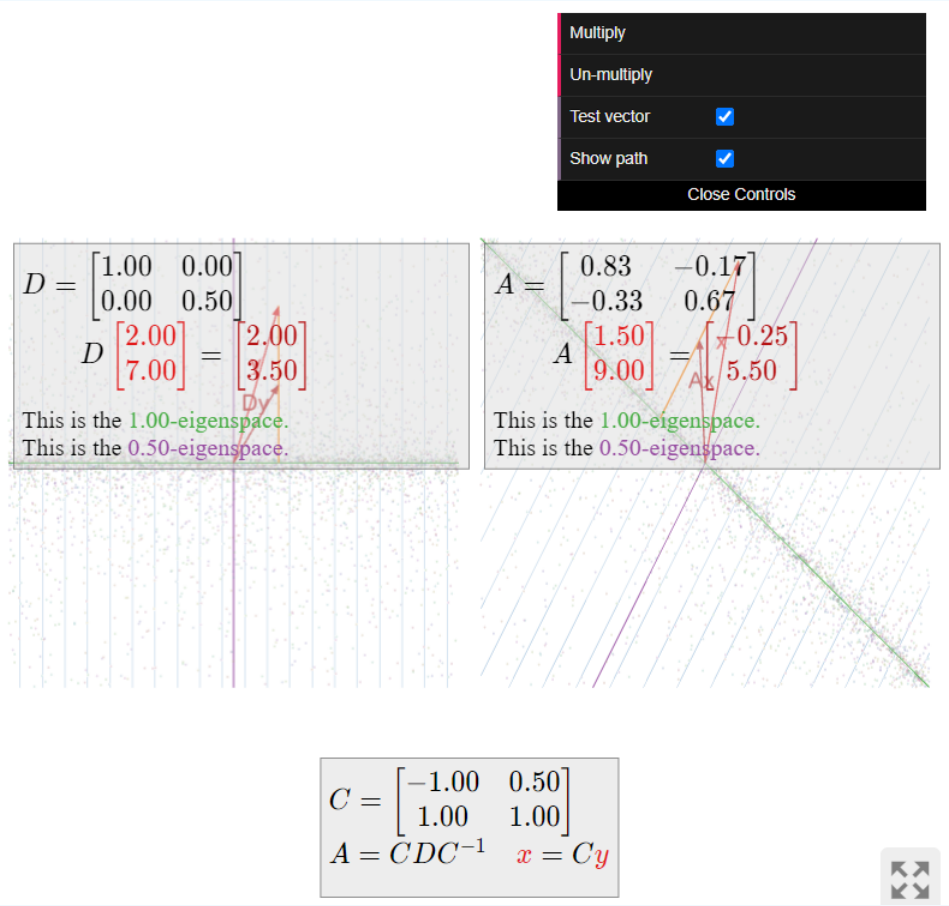


Figure 6.3.11: Dynamics of the matrices A and D. Click "multiply" to multiply the colored points by D on the left and A on the right.

Example 6.3.17: Interactive: A diagonalizable 3×3 matrix

The diagonal matrix

$$D = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3/2 \end{pmatrix}$$

scales the x-coordinate by $1/2$, the y-coordinate by 2, and the z-coordinate by $3/2$. Looking straight down at the xy-plane, the points follow parabolic paths taking them away from the x-axis and toward the y-axis. The z-coordinate is scaled by $3/2$, so points fly away from the xy-plane in that direction.

If $A = CDC^{-1}$ for some invertible matrix C, then A does the same thing as D, but with respect to the coordinate system defined by the columns of C.



Figure 6.3.12: Dynamics of the matrices A and D. Click “multiply” to multiply the colored points by D on the left and A on the right.

Algebraic and Geometric Multiplicity

In this subsection, we give a variant of Theorem 6.3.1 that provides another criterion for diagonalizability. It is stated in the language of *multiplicities* of eigenvalues.

In algebra, we define the *multiplicity* of a root λ_0 of a polynomial $f(\lambda)$ to be the number of factors of $\lambda - \lambda_0$ that divide $f(\lambda)$. For instance, in the polynomial

$$f(\lambda) = -\lambda^3 + 4\lambda^2 - 5\lambda + 2 = -(\lambda - 1)^2(\lambda - 2),$$

the root $\lambda_0 = 2$ has multiplicity 1, and the root $\lambda_0 = 1$ has multiplicity 2.

Definition 6.3.2: Algebraic Multiplicity and Geometric Multiplicity

Let A be an $n \times n$ matrix, and let λ be an eigenvalue of A.

1. The ***algebraic multiplicity*** of λ is its multiplicity as a root of the characteristic polynomial of A.
2. The ***geometric multiplicity*** of λ is the dimension of the λ-eigenspace.

Since the λ-eigenspace of A is $\text{Nul}(A - \lambda I_n)$, its dimension is the number of free variables in the system of equations $(A - \lambda I_n)x = 0$, i.e., the number of columns without pivots in the matrix $A - \lambda I_n$.

Example 6.3.18

The shear matrix

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

has only one eigenvalue $\lambda = 1$. The characteristic polynomial of A is $f(\lambda) = (\lambda - 1)^2$, so 1 has algebraic multiplicity 2, as it is a double root of f. On the other hand, we showed in Example 6.3.9 that the 1-eigenspace of A is the x-axis, so the geometric multiplicity of 1 is equal to 1. This matrix is not diagonalizable.

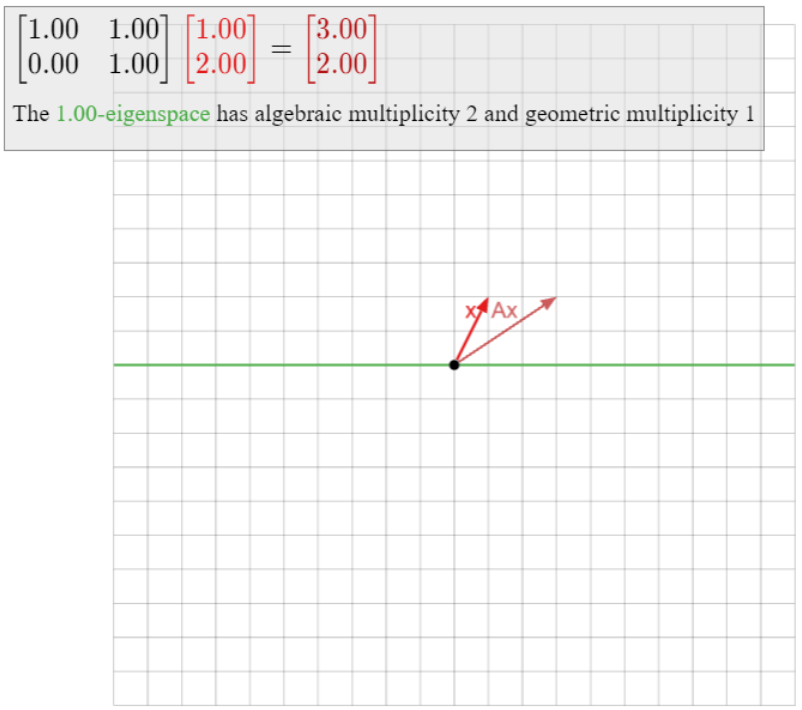


Figure 6.3.13: Eigenspace of the shear matrix, with multiplicities.

The identity matrix

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

also has characteristic polynomial $(\lambda - 1)^2$, so the eigenvalue 1 has algebraic multiplicity 2. Since every nonzero vector in $\mathbb{R}^2$ is an eigenvector of I_2 with eigenvalue 1, the 1-eigenspace is all of $\mathbb{R}^2$, so the geometric multiplicity is 2 as well. This matrix is diagonal.

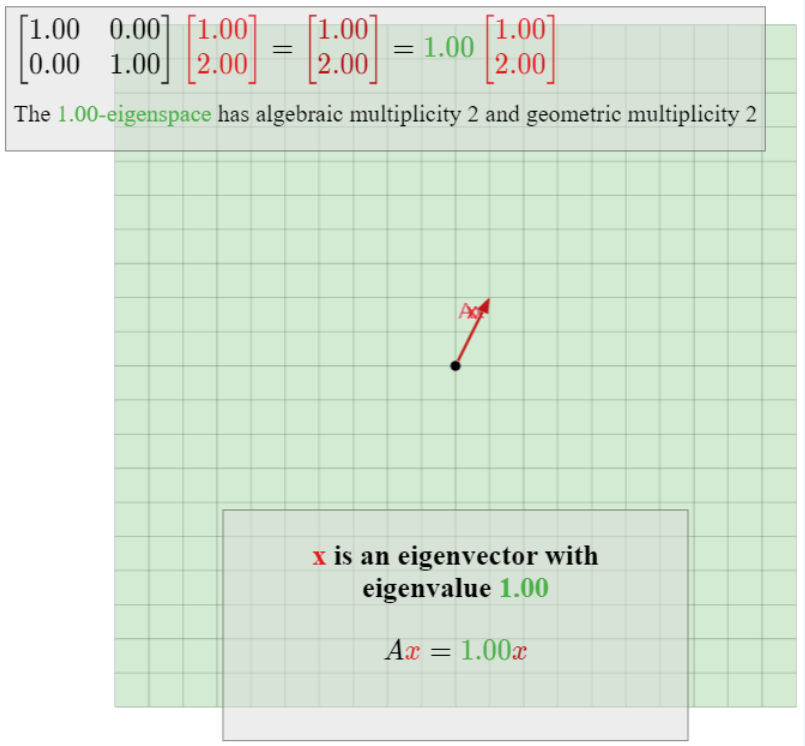


Figure 6.3.14: Eigenspace of the identity matrix, with multiplicities.

Example 6.3.19

Continuing with Example 6.3.8, let

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

The characteristic polynomial is $f(\lambda) = -(\lambda-1)^2(\lambda-2)$, so that 1 and 2 are the eigenvalues, with algebraic multiplicities 2 and 1, respectively. We computed that the 1-eigenspace is a plane and the 2-eigenspace is a line, so that 1 and 2 also have geometric multiplicities 2 and 1, respectively. This matrix is diagonalizable.

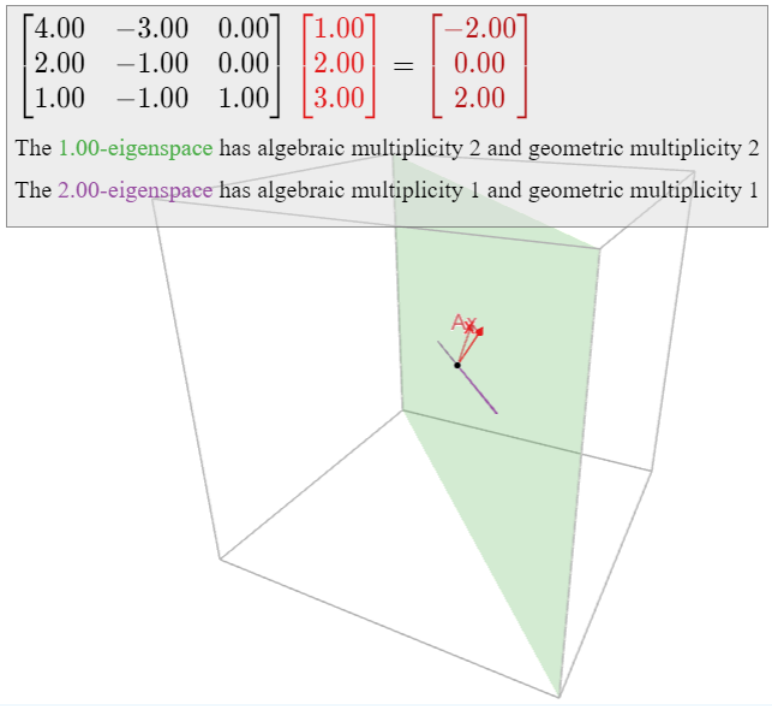


Figure 6.3.15: The green plane is the 1-eigenspace of A, and the violet line is the 2-eigenspace. Hence the geometric multiplicity of the 1-eigenspace is 2, and the geometric multiplicity of the 2-eigenspace is 1.

In Example 6.3.11, we saw that the matrix

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

also has characteristic polynomial $f(\lambda) = -(\lambda - 1)^2(\lambda - 2)$, so that 1 and 2 are the eigenvalues, with algebraic multiplicities 2 and 1, respectively. In this case, however, both eigenspaces are *lines*, so that both eigenvalues have geometric multiplicity 1. This matrix is not diagonalizable.

$$\begin{bmatrix} 1.00 & 1.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 2.00 \end{bmatrix} \begin{bmatrix} 1.00 \\ 2.00 \\ 3.00 \end{bmatrix} = \begin{bmatrix} 3.00 \\ 2.00 \\ 6.00 \end{bmatrix}$$

The 1.00-eigenspace has algebraic multiplicity 2 and geometric multiplicity 1

The 2.00-eigenspace has algebraic multiplicity 1 and geometric multiplicity 1

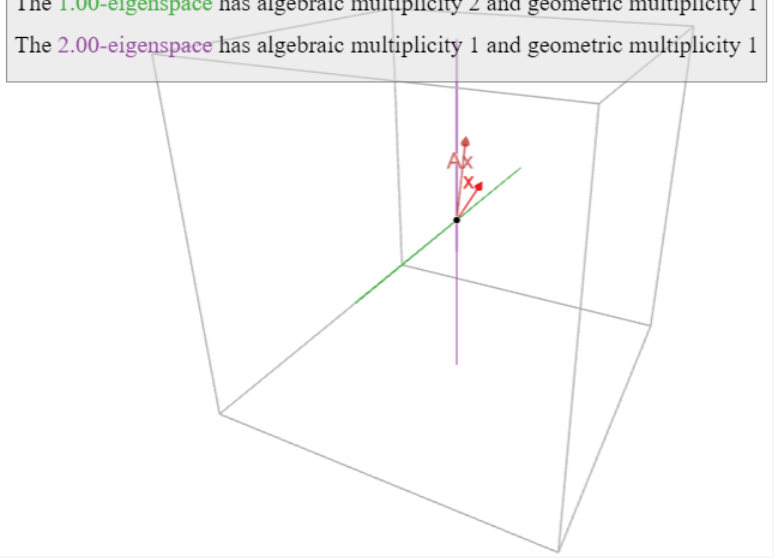


Figure 6.3.16: Both eigenspaces of A are lines, so they both have geometric multiplicity 1.

We saw in the above examples that the algebraic and geometric multiplicities need not coincide. However, they do satisfy the following fundamental inequality, the proof of which is beyond the scope of this text.

Theorem 6.3.2: Algebraic and Geometric Multiplicity

Let A be a square matrix and let λ be an eigenvalue of A. Then

$$1 \leq (\text{the geometric multiplicity of } \lambda) \leq (\text{the algebraic multiplicity of } \lambda).$$

In particular, if the algebraic multiplicity of λ is equal to 1, then so is the geometric multiplicity.

Note 6.3.4

If A has an eigenvalue λ with algebraic multiplicity 1, then the λ-eigenspace is a *line*.

We can use Theorem 6.3.2 to give another criterion for diagonalizability (in addition to Theorem 6.3.1).

Theorem 6.3.3: Diagonalization Theorem, Variant

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

1. A is diagonalizable.
2. The sum of the geometric multiplicities of the eigenvalues of A is equal to n.
3. The sum of the algebraic multiplicities of the eigenvalues of A is equal to n, and for each eigenvalue, the geometric multiplicity equals the algebraic multiplicity.

Proof

We will show $1 \implies 2 \implies 3 \implies 1$. First suppose that A is diagonalizable. Then A has n linearly independent eigenvectors $v_1, v_2, \ldots, v_n$. This implies that the sum of the geometric multiplicities is *at least* n: for instance, if v_1, v_2, v_3 have the same eigenvalue λ, then the geometric multiplicity of λ is at least 3 (as the λ-eigenspace contains three linearly independent vectors), and so on. But the sum of the algebraic multiplicities is greater than or equal to the sum of the

geometric multiplicities by Theorem 6.3.2 and the sum of the algebraic multiplicities is at most n because the characteristic polynomial has degree n. Therefore, the sum of the geometric multiplicities equals n.

Now suppose that the sum of the geometric multiplicities equals n. As above, this forces the sum of the algebraic multiplicities to equal n as well. As the algebraic multiplicities are all greater than or equal to the geometric multiplicities in any case, this implies that they are in fact equal.

Finally, suppose that the third condition is satisfied. Then the sum of the geometric multiplicities equals n. Suppose that the distinct eigenvectors are $\lambda_1, \lambda_2, \ldots, \lambda_k$, and that $\mathcal{B}_i$ is a basis for the λ_i-eigenspace, which we call V_i. We claim that the collection $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ of all vectors in all of the eigenspace bases $\mathcal{B}_i$ is linearly independent. Consider the vector equation

$$0 = c_1v_1 + c_2v_2 + \cdots + c_nv_n.$$

Grouping the eigenvectors with the same eigenvalues, this sum has the form

$$0 = (\text{something in } V_1) + (\text{something in } V_2) + \cdots + (\text{something in } V_k).$$

Since eigenvectors with distinct eigenvalues are linearly independent, Fact 5.1.1 in Section 5.1, each "something in V_i" is equal to zero. But this implies that all coefficients $c_1, c_2, \ldots, c_n$ are equal to zero, since the vectors in each $\mathcal{B}_i$ are linearly independent. Therefore, A has n linearly independent eigenvectors, so it is diagonalizable.

The first part of the third statement simply says that the characteristic polynomial of A factors completely into linear polynomials over the real numbers: in other words, there are no complex (non-real) roots. The second part of the third statement says in particular that for any diagonalizable matrix, the algebraic and geometric multiplicities coincide.

Note 6.3.5

Let A be a square matrix and let λ be an eigenvalue of A. If the algebraic multiplicity of λ does not equal the geometric multiplicity, then A is not diagonalizable.

The examples at the beginning of this subsection illustrate the theorem. Here we give some general consequences for diagonalizability of 2×2 and 3×3 matrices.

Example 6.3.20: Diagonalizability of 2×2 Matrices

Let A be a 2×2 matrix. There are four cases:

1. A has two different eigenvalues. In this case, each eigenvalue has algebraic and geometric multiplicity equal to one. This implies A is diagonalizable. For example:
$$A = \begin{pmatrix} 1 & 7 \\ 0 & 2 \end{pmatrix}.$$
2. A has one eigenvalue λ of algebraic and geometric multiplicity 2. To say that the geometric multiplicity is 2 means that $\text{Nul}(A - \lambda I_2) = \mathbb{R}^2$, i.e., that every vector in $\mathbb{R}^2$ is in the null space of $A - \lambda I_2$. This implies that $A - \lambda I_2$ is the zero matrix, so that A is the diagonal matrix λI_2. In particular, A is diagonalizable. For example:
$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$
3. A has one eigenvalue λ of algebraic multiplicity 2 and geometric multiplicity 1. In this case, A is not diagonalizable, by part 3 of Theorem 6.3.3. For example, a shear:
$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$
4. A has no eigenvalues. This happens when the characteristic polynomial has no real roots. In particular, A is not diagonalizable. For example, a rotation:

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Example 6.3.21: Diagonalizability of 3×3 Matrices

Let A be a 3×3 matrix. We can analyze the diagonalizability of A on a case-by-case basis, as in the previous Example 6.3.20

1. A has three different eigenvalues. In this case, each eigenvalue has algebraic and geometric multiplicity equal to one. This implies A is diagonalizable. For example:
$$A = \begin{pmatrix} 1 & 7 & 4 \\ 0 & 2 & 3 \\ 0 & 0 & -1 \end{pmatrix}.$$
2. A has two distinct eigenvalues λ_1, λ_2. In this case, one has algebraic multiplicity one and the other has algebraic multiplicity two; after reordering, we can assume λ_1 has multiplicity 1 and λ_2 has multiplicity 2. This implies that λ_1 has geometric multiplicity 1, so *A is diagonalizable if and only if the λ_2-eigenspace is a plane.* For example:
$$A = \begin{pmatrix} 1 & 7 & 4 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$
On the other hand, if the geometric multiplicity of λ_2 is 1, then A is not diagonalizable. For example:
$$A = \begin{pmatrix} 1 & 7 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$
3. A has only one eigenvalue λ. If the algebraic multiplicity of λ is 1, then A is not diagonalizable. This happens when the characteristic polynomial has two complex (non-real) roots. For example:
$$A = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$
Otherwise, the algebraic multiplicity of λ is equal to 3. In this case, if the geometric multiplicity is 1:
$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$
or 2:
$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$
then A is not diagonalizable. If the geometric multiplicity is 3, then $\text{Nul}(A - \lambda I_3) = \mathbb{R}^3$, so that $A - \lambda I_3$ is the zero matrix, and hence $A = \lambda I_3$. Therefore, in this case A is necessarily diagonal, as in:
$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Similarity and Multiplicity

Recall from Fact 5.3.2 in Section 5.3 that similar matrices have the same eigenvalues. It turns out that *both* notions of multiplicity of an eigenvalue are preserved under similarity.

Theorem 6.3.4

Let A and B be similar $n \times n$ matrices, and let λ be an eigenvalue of A and B. Then:

1. The algebraic multiplicity of λ is the same for A and B.
2. The geometric multiplicity of λ is the same for A and B.

Proof

Since A and B have the same characteristic polynomial, the multiplicity of λ as a root of the characteristic polynomial is the same for both matrices, which proves the first statement. For the second, suppose that $A = CBC^{-1}$ for an invertible matrix C. By Fact 5.3.3 in Section 5.3, the matrix C takes eigenvectors of B to eigenvectors of A, both with eigenvalue λ.

Let $\{v_1, v_2, \ldots, v_k\}$ be a basis of the λ-eigenspace of B. We claim that $\{Cv_1, Cv_2, \ldots, Cv_k\}$ is linearly independent. Suppose that

$$c_1Cv_1 + c_2Cv_2 + \cdots + c_kCv_k = 0.$$

Regrouping, this means

$$C\big(c_1v_1 + c_2v_2 + \cdots + c_kv_k\big) = 0.$$

By Theorem 5.1.1 in Section 5.1, the null space of C is trivial, so this implies

$$c_1v_1 + c_2v_2 + \cdots + c_kv_k = 0.$$

Since $v_1, v_2, \ldots, v_k$ are linearly independent, we get $c_1 = c_2 = \cdots = c_k = 0$, as desired.

By the previous paragraph, the dimension of the λ-eigenspace of A is greater than or equal to the dimension of the λ-eigenspace of B. By symmetry (B is similar to A as well), the dimensions are equal, so the geometric multiplicities coincide.

For instance, the four matrices in Example 6.3.20 are not similar to each other, because the algebraic and/or geometric multiplicities of the eigenvalues do not match up. Or, combined with the above Theorem 6.3.3, we see that a diagonalizable matrix cannot be similar to a non-diagonalizable one, because the algebraic and geometric multiplicities of such matrices cannot both coincide.

Example 6.3.22

Continuing with this Example 6.3.8, let

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

This is a diagonalizable matrix that is similar to

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \quad \text{using the matrix} \quad C = \begin{pmatrix} 1 & 0 & 3 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}.$$

The 1-eigenspace of D is the xy-plane, and the 2-eigenspace is the z-axis. The matrix C takes the xy-plane to the 1-eigenspace of A, which is again a plane, and the z-axis to the 2-eigenspace of A, which is again a line. This shows that the geometric multiplicities of A and D coincide.

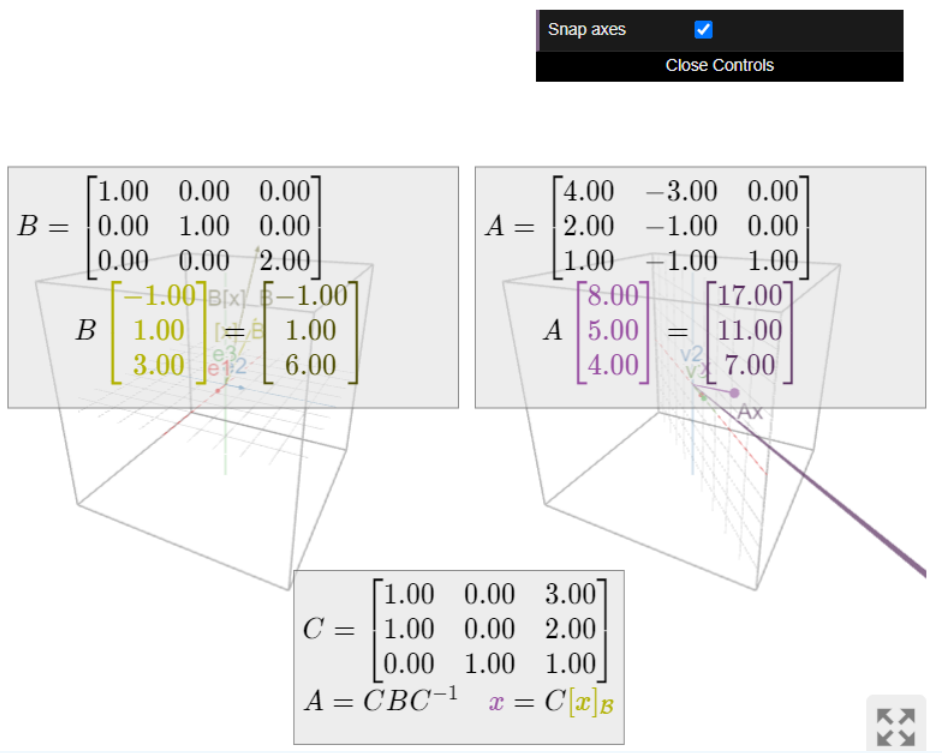


Figure 6.3.17: The matrix C takes the xy-plane to the 1-eigenspace of A (the grid) and the z-axis to the 2-eigenspace (the green line).

The converse of Theorem 6.3.4 is false: there exist matrices whose eigenvectors have the same algebraic and geometric multiplicities, but which are not similar. See Example 6.3.23 below. However, for 2×2 and 3×3 matrices whose characteristic polynomial has no complex (non-real) roots, the converse of Theorem 6.3.4 is true. (We will handle the case of complex roots in Section 5.5.)

Example 6.3.23: Matrices that look similar but are not

Show that the matrices

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

have the same eigenvalues with the same algebraic and geometric multiplicities, but are not similar.

Solution

These matrices are upper-triangular. They both have characteristic polynomial $f(\lambda) = \lambda^4$, so they both have one eigenvalue 0 with algebraic multiplicity 4. The 0-eigenspace is the null space, Fact 5.1.2 in Section 5.1, which has dimension 2 in each case because A and B have two columns without pivots. Hence 0 has geometric multiplicity 2 in each case.

To show that A and B are not similar, we note that

$$A^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad B^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

as the reader can verify. If $A = CBC^{-1}$ then by Recipe: Compute Powers of a Diagonalizable Matrix, we have

$$A^2 = CB^2C^{-1} = C0C^{-1} = 0,$$

which is not the case.

On the other hand, suppose that A and B are *diagonalizable* matrices with the same characteristic polynomial. Since the geometric multiplicities of the eigenvalues coincide with the algebraic multiplicities, which are the same for A and B, we conclude that there exist n linearly independent eigenvectors of each matrix, all of which have the same eigenvalues. This shows that A and B are both similar to the same diagonal matrix. Using the transitivity property, Proposition 5.3.1 in Section 5.3, of similar matrices, this shows:

Note 6.3.6

Diagonalizable matrices are similar if and only if they have the same characteristic polynomial, or equivalently, the same eigenvalues with the same algebraic multiplicities.

Example 6.3.24

Show that the matrices

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

are similar.

Solution

Both matrices have the three distinct eigenvalues $1, -1, 4$. Hence they are both diagonalizable, and are similar to the diagonal matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

By the transitivity property, Proposition 5.3.1 in Section 5.3, of similar matrices, this implies that A and B are similar to each other.

Example 6.3.25: Diagonal matrices with the same entries are similar

Any two diagonal matrices with the same diagonal entries (possibly in a different order) are similar to each other. Indeed, such matrices have the same characteristic polynomial. We saw this phenomenon in Example 6.3.5, where we noted that

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}^{-1}.$$

This page titled 6.3: Diagonalization is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **5.4: Diagonalization** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

6.4: Complex Eigenvalues

Objectives

1. Learn to find complex eigenvalues and eigenvectors of a matrix.
2. Learn to recognize a rotation-scaling matrix, and compute by how much the matrix rotates and scales.
3. Understand the geometry of 2×2 and 3×3 matrices with a complex eigenvalue.
4. *Recipes:* a 2×2 matrix with a complex eigenvalue is similar to a rotation-scaling matrix, the eigenvector trick for 2×2 matrices.
5. *Pictures:* the geometry of matrices with a complex eigenvalue.
6. *Theorems:* the rotation-scaling theorem, the block diagonalization theorem.
7. *Vocabulary word:* ***rotation-scaling matrix***.

In Section 5.4, we saw that an $n \times n$ matrix whose characteristic polynomial has n distinct real roots is *diagonalizable*: it is similar to a diagonal matrix, which is much simpler to analyze. The other possibility is that a matrix has *complex* roots, and that is the focus of this section. It turns out that such a matrix is similar (in the 2×2 case) to a *rotation-scaling matrix*, which is also relatively easy to understand.

In a certain sense, this entire section is analogous to Section 5.4, with rotation-scaling matrices playing the role of diagonal matrices. See Section 7.1 for a review of the complex numbers.

Matrices with Complex Eigenvalues

As a consequence of the fundamental theorem of algebra, Theorem 7.1.1 in Section 7.1, as applied to the characteristic polynomial, we see that:

Note 6.4.1

Every $n \times n$ matrix has exactly n complex eigenvalues, counted with multiplicity.

We can compute a corresponding (complex) eigenvector in exactly the same way as before: by row reducing the matrix $A - \lambda I_n$. Now, however, we have to do arithmetic with complex numbers.

Example 6.4.1: A 2×2 matrix

Find the complex eigenvalues and eigenvectors of the matrix

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Solution

The characteristic polynomial of A is

$$f(\lambda) = \lambda^2 - \operatorname{Tr}(A)\lambda + \det(A) = \lambda^2 - 2\lambda + 2.$$

The roots of this polynomial are

$$\lambda = \frac{2 \pm \sqrt{4-8}}{2} = 1 \pm i.$$

First we compute an eigenvector for $\lambda = 1+i$. We have

$$A - (1+i)I_2 = \begin{pmatrix} 1-(1+i) & -1 \\ 1 & 1-(1+i) \end{pmatrix} = \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix}.$$

Now we row reduce, noting that the second row is i times the first:

$$\begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} \xrightarrow{R_2 = R_2 - iR_1} \begin{pmatrix} -i & -1 \\ 0 & 0 \end{pmatrix} \xrightarrow{R_1 = R_1 \div -i} \begin{pmatrix} 1 & -i \\ 0 & 0 \end{pmatrix}.$$

The parametric form is $x = iy$, so that an eigenvector is $v_1 = \begin{pmatrix} i \\ 1 \end{pmatrix}$. Next we compute an eigenvector for $\lambda = 1 - i$. We have

$$A - (1-i)I_2 = \begin{pmatrix} 1-(1-i) & -1 \\ 1 & 1-(1-i) \end{pmatrix} = \begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix}.$$

Now we row reduce, noting that the second row is $-i$ times the first:

$$\begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix} \xrightarrow{R_2 = R_2 + iR_1} \begin{pmatrix} i & -1 \\ 0 & 0 \end{pmatrix} \xrightarrow{R_1 = R_1 \div i} \begin{pmatrix} 1 & i \\ 0 & 0 \end{pmatrix}.$$

The parametric form is $x = -iy$, so that an eigenvector is $v_2 = \begin{pmatrix} -i \\ 1 \end{pmatrix}$.

We can verify our answers:

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} i \\ 1 \end{pmatrix} = \begin{pmatrix} i-1 \\ i+1 \end{pmatrix} = (1+i) \begin{pmatrix} i \\ 1 \end{pmatrix}$$
$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -i \\ 1 \end{pmatrix} = \begin{pmatrix} -i-1 \\ -i+1 \end{pmatrix} = (1-i) \begin{pmatrix} -i \\ 1 \end{pmatrix}.$$

Example 6.4.2: A 3×3 matrix

Find the eigenvalues and eigenvectors, real and complex, of the matrix

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

Solution

We compute the characteristic polynomial by expanding cofactors along the third row:

$$f(\lambda) = \det \begin{pmatrix} 4/5 - \lambda & -3/5 & 0 \\ 3/5 & 4-5-\lambda & 0 \\ 1 & 2 & 2-\lambda \end{pmatrix} = (2-\lambda)\left(\lambda^2 - \frac{8}{5}\lambda + 1\right).$$

This polynomial has one real root at 2, and two complex roots at

$$\lambda = \frac{8/5 \pm \sqrt{64/25 - 4}}{2} = \frac{4 \pm 3i}{5}.$$

Therefore, the eigenvalues are

$$\lambda = 2, \quad \frac{4+3i}{5}, \quad \frac{4-3i}{5}.$$

We eyeball that $v_1 = e_3$ is an eigenvector with eigenvalue 2, since the third column is $2e_3$.

Next we find an eigenvector with eigenvaluue $(4+3i)/5$. We have

$$A - \frac{4+3i}{5} I_3 = \begin{pmatrix} -3i/5 & -3/5 & 0 \\ 3/5 & -3i/5 & 0 \\ 1 & 2 & 2-(4+3i)/5 \end{pmatrix} \xrightarrow[R_2 = R_2 \times 5/3]{R_1 = R_1 \times -5/3} \begin{pmatrix} i & 1 & 0 \\ 1 & -i & 0 \\ 1 & 2 & \frac{6-3i}{5} \end{pmatrix}.$$

We row reduce, noting that the second row is $-i$ times the first:

$$\begin{pmatrix} i & 1 & 0 \\ 1 & -i & 0 \\ 1 & 2 & \frac{6-3i}{5} \end{pmatrix} \xrightarrow{R_2=R_2+iR_1} \begin{pmatrix} i & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 2 & \frac{6-3i}{5} \end{pmatrix}$$
$$\xrightarrow{R_3=R_3+iR_1} \begin{pmatrix} i & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 2+i & \frac{6-3i}{5} \end{pmatrix}$$
$$\xrightarrow{R_2\longleftrightarrow R_3} \begin{pmatrix} i & 1 & 0 \\ 0 & 2+i & \frac{6-3i}{5} \\ 0 & 0 & 0 \end{pmatrix}$$
$$\xrightarrow[R_2=R_2\div(2+i)]{R_1=R_1\div i} \begin{pmatrix} 1 & -i & 0 \\ 0 & 1 & \frac{9-12i}{25} \\ 0 & 0 & 0 \end{pmatrix}$$
$$\xrightarrow{R_1=R_1+iR_2} \begin{pmatrix} 1 & 0 & \frac{12+9i}{25} \\ 0 & 1 & \frac{9-12i}{25} \\ 0 & 0 & 0 \end{pmatrix}.$$

The free variable is z; the parametric form of the solution is

$$\begin{cases} x &= -\dfrac{12+9i}{25}z \\ y &= -\dfrac{9-12i}{25}z. \end{cases}$$

Taking $z = 25$ gives the eigenvector

$$v_2 = \begin{pmatrix} -12-9i \\ -9+12i \\ 25 \end{pmatrix}.$$

A similar calculation (replacing all occurences of i by $-i$) shows that an eigenvector with eigenvalue $(4-3i)/5$ is

$$v_3 = \begin{pmatrix} -12+9i \\ -9-12i \\ 25 \end{pmatrix}.$$

We can verify our calculations:

$$\begin{pmatrix} 4/5 & -3/5 & 0 \\ 3/5 & 4/5 & 0 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} -12+9i \\ -9-12i \\ 25 \end{pmatrix} = \begin{pmatrix} -21/5+72i/5 \\ -72/5-21i/5 \\ 20-15i \end{pmatrix} = \frac{4+3i}{5} \begin{pmatrix} -12+9i \\ -9-12i \\ 25 \end{pmatrix}$$
$$\begin{pmatrix} 4/5 & -3/5 & 0 \\ 3/5 & 4/5 & 0 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} -12-9i \\ -9+12i \\ 25 \end{pmatrix} = \begin{pmatrix} -21/5-72i/5 \\ -72/5+21i/5 \\ 20+15i \end{pmatrix} = \frac{4-3i}{5} \begin{pmatrix} -12-9i \\ -9+12i \\ 25 \end{pmatrix}.$$

If A is a matrix with real entries, then its characteristic polynomial has real coefficients, so Note 7.1.3 in Section 7.1 implies that its complex eigenvalues come in conjugate pairs. In the first example, we notice that

$$1+i \text{ has an eigenvector } v_1 = \begin{pmatrix} i \\ 1 \end{pmatrix}$$
$$1-i \text{ has an eigenvector } v_2 = \begin{pmatrix} -i \\ 1 \end{pmatrix}.$$

In the second example,

$$\frac{4+3i}{5} \text{ has an eigenvector } v_1 = \begin{pmatrix} -12-9i \\ -9+12i \\ 25 \end{pmatrix}$$
$$\frac{4-3i}{5} \text{ has an eigenvector } v_2 = \begin{pmatrix} -12+9i \\ -9-12i \\ 25 \end{pmatrix}$$

In these cases, an eigenvector for the conjugate eigenvalue is simply the *conjugate eigenvector* (the eigenvector obtained by conjugating each entry of the first eigenvector). This is always true. Indeed, if $Av = \lambda v$ then

$$A\bar{v} = \bar{A}\bar{v} = \bar{\lambda}\bar{v} = \bar{\lambda}\bar{v},$$

which exactly says that $\bar{v}$ is an eigenvector of A with eigenvalue $\bar{\lambda}$.

Note 6.4.2

Let A be a matrix with real entries. If

$$\begin{gathered} \lambda \text{ is a complex eigenvalue with eigenvector } v, \\ \text{then } \bar{\lambda} \text{ is a complex eigenvalue with eigenvector } \bar{v}. \end{gathered}$$

In other words, both eigenvalues and eigenvectors come in conjugate pairs.

Since it can be tedious to divide by complex numbers while row reducing, it is useful to learn the following trick, which works equally well for matrices with real entries.

Note 6.4.3: Eigenvector Trick for 2×2 Matrices

Let A be a 2×2 matrix, and let λ be a (real or complex) eigenvalue. Then

$$A - \lambda I_2 = \begin{pmatrix} z & w \\ \star & \star \end{pmatrix} \quad \Longrightarrow \quad \begin{pmatrix} -w \\ z \end{pmatrix} \text{ is an eigenvector with eigenvalue } \lambda,$$

assuming the first row of $A - \lambda I_2$ is nonzero.

Indeed, since λ is an eigenvalue, we know that $A - \lambda I_2$ is not an invertible matrix. It follows that the rows are collinear (otherwise the determinant is nonzero), so that the second row is automatically a (complex) multiple of the first:

$$\begin{pmatrix} z & w \\ \star & \star \end{pmatrix} = \begin{pmatrix} z & w \\ cz & cw \end{pmatrix}.$$

It is obvious that $\begin{pmatrix} -w \\ z \end{pmatrix}$ is in the null space of this matrix, as is $\begin{pmatrix} w \\ -z \end{pmatrix}$, for that matter. Note that we never had to compute the second row of $A - \lambda I_2$, let alone row reduce!

Example 6.4.3: A 2×2 matrix, the easy way

Find the complex eigenvalues and eigenvectors of the matrix

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Solution

Since the characteristic polynomial of a 2×2 matrix A is $f(\lambda) = \lambda^2 - \operatorname{Tr}(A)\lambda + \det(A)$, its roots are

$$\lambda = \frac{\operatorname{Tr}(A) \pm \sqrt{\operatorname{Tr}(A)^2 - 4\det(A)}}{2} = \frac{2 \pm \sqrt{4-8}}{2} = 1 \pm i.$$

To find an eigenvector with eigenvalue $1+i$, we compute

$$A-(1+i)I_2=\begin{pmatrix}-i & -1\\ \star & \star\end{pmatrix}\xrightarrow{\text{eigenvector}} v_1=\begin{pmatrix}1\\ -i\end{pmatrix}.$$

The eigenvector for the conjugate eigenvalue is the complex conjugate:

$$v_2=\bar{v}_1=\begin{pmatrix}1\\ i\end{pmatrix}.$$

In Example 6.4.1 we found the eigenvectors $\begin{pmatrix}i\\1\end{pmatrix}$ and $\begin{pmatrix}-i\\1\end{pmatrix}$ for the eigenvalues $1+i$ and $1-i$, respectively, but in Example 6.4.3 we found the eigenvectors $\begin{pmatrix}1\\-i\end{pmatrix}$ and $\begin{pmatrix}1\\i\end{pmatrix}$ for the same eigenvalues of the same matrix. These vectors do not look like multiples of each other at first—but since we now have complex numbers at our disposal, we can see that they actually are multiples:

$$-i\begin{pmatrix}i\\1\end{pmatrix}=\begin{pmatrix}1\\-i\end{pmatrix}\qquad i\begin{pmatrix}-i\\1\end{pmatrix}=\begin{pmatrix}1\\i\end{pmatrix}.$$

Rotation-Scaling Matrices

The most important examples of matrices with complex eigenvalues are rotation-scaling matrices, i.e., scalar multiples of rotation matrices.

Definition 6.4.1: Rotation-Scaling matrix

A ***rotation-scaling matrix*** is a 2×2 matrix of the form

$$\begin{pmatrix}a & -b\\ b & a\end{pmatrix},$$

where a and b are real numbers, not both equal to zero.

The following proposition justifies the name.

Proposition 6.4.1

Let

$$A=\begin{pmatrix}a & -b\\ b & a\end{pmatrix}$$

be a rotation-scaling matrix. Then:

1. A is a product of a rotation matrix

$$\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\quad\text{with a scaling matrix}\quad\begin{pmatrix}r & 0\\ 0 & r\end{pmatrix}.$$

2. The scaling factor r is

$$r=\sqrt{\det(A)}=\sqrt{a^2+b^2}.$$

3. The rotation angle θ is the counterclockwise angle from the positive x-axis to the vector $\begin{pmatrix}a\\b\end{pmatrix}$:

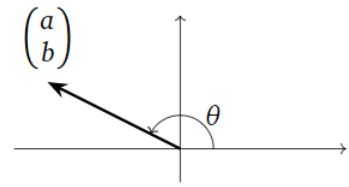


Figure 6.4.1

The eigenvalues of A are $\lambda=a\pm bi$.

Proof

Set $r = \sqrt{\det(A)} = \sqrt{a^2+b^2}$. The point $(a/r, b/r)$ has the property that

$$\left(\frac{a}{r}\right)^2 + \left(\frac{b}{r}\right)^2 = \frac{a^2+b^2}{r^2} = 1.$$

In other words $(a/r, b/r)$ lies on the unit circle. Therefore, it has the form $(\cos\theta, \sin\theta)$, where θ is the counterclockwise angle from the positive x-axis to the vector $\begin{pmatrix} a/r \\ b/r \end{pmatrix}$, or since it is on the same line, to $\begin{pmatrix} a \\ b \end{pmatrix}$:

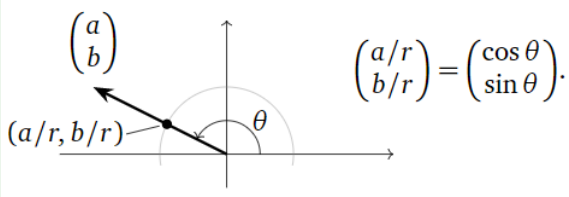


Figure 6.4.2

It follows that

$$A = r\begin{pmatrix} a/r & -b/r \\ b/r & a/r \end{pmatrix} = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

as desired.

For the last statement, we compute the eigenvalues of A as the roots of the characteristic polynomial:

$$\lambda = \frac{\operatorname{Tr}(A) \pm \sqrt{\operatorname{Tr}(A)^2 - 4\det(A)}}{2} = \frac{2a \pm \sqrt{4a^2 - 4(a^2+b^2)}}{2} = a \pm bi.$$

Geometrically, a rotation-scaling matrix does exactly what the name says: it rotates and scales (in either order).

Example 6.4.4: A rotation-scaling matrix

What does the matrix

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

do geometrically?

Solution

This is a rotation-scaling matrix with $a = b = 1$. Therefore, it scales by a factor of $\sqrt{\det(A)} = \sqrt{2}$ and rotates counterclockwise by 45°:

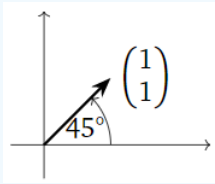


Figure 6.4.3

Here is a picture of A:

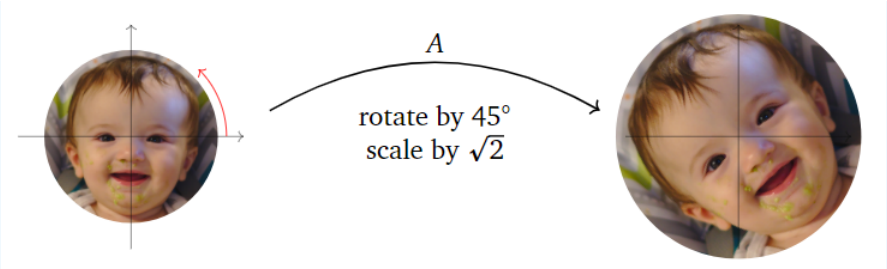


Figure 6.4.4

An interactive figure is included below.

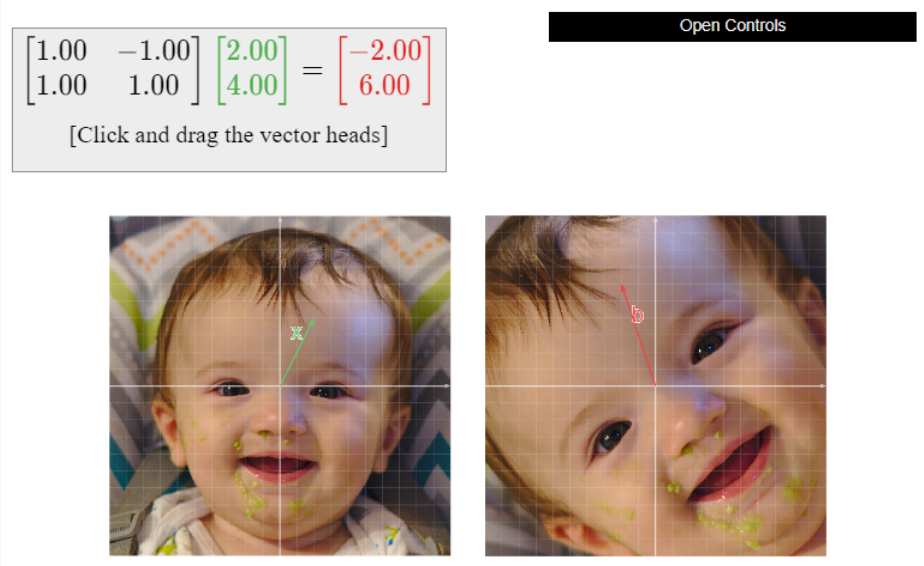


Figure 6.4.5: Multiplication by the matrix A rotates the plane by 45° and dilates by a factor of $\sqrt{2}$. Move the input vector x to see how the output vector b changes.

Example 6.4.5: A rotation-scaling matrix

What does the matrix

$$A = \begin{pmatrix} -\sqrt{3} & -1 \\ 1 & -\sqrt{3} \end{pmatrix}$$

do geometrically?

Solution

This is a rotation-scaling matrix with $a = -\sqrt{3}$ and $b = 1$. Therefore, it scales by a factor of $\sqrt{\det(A)} = \sqrt{3+1} = 2$ and rotates counterclockwise by the angle θ in the picture:

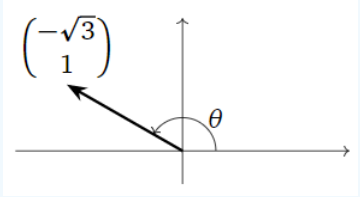


Figure 6.4.6

To compute this angle, we do a bit of trigonometry:

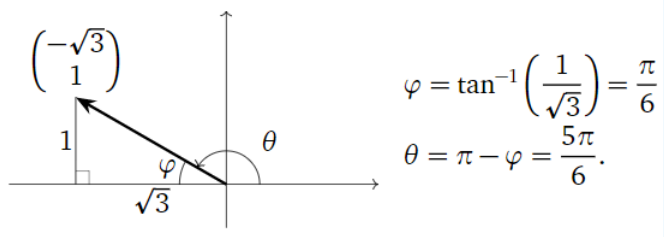


Figure 6.4.7

Therefore, A rotates counterclockwise by $5\pi/6$ and scales by a factor of 2.

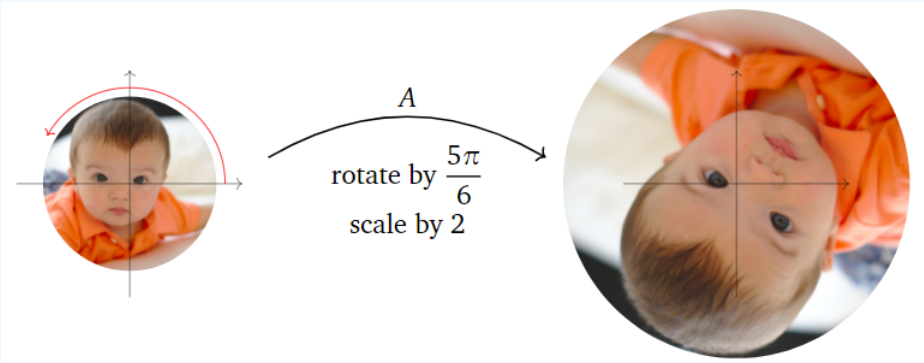


Figure 6.4.8

An interactive figure is included below.

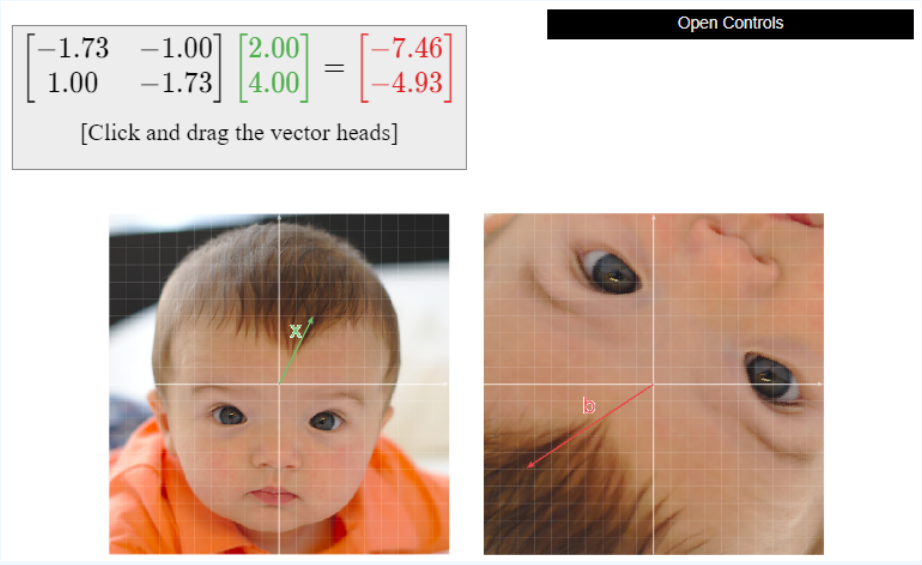
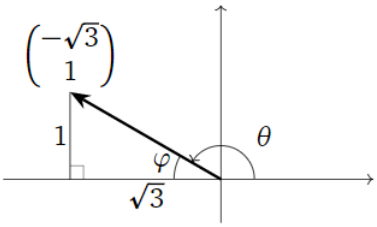


Figure 6.4.9: Multiplication by the matrix A rotates the plane by $5\pi/6$ and dilates by a factor of 2. Move the input vector x to see how the output vector b changes.

The matrix in the second example has second column $\begin{pmatrix}-\sqrt{3}\\1\end{pmatrix}$, which is rotated counterclockwise from the positive x-axis by an angle of $5\pi/6$. This rotation angle is *not* equal to $\tan^{-1}\left(1/(-\sqrt{3})\right) = -\frac{\pi}{6}$. The problem is that arctan always outputs values between $-\pi/2$ and $\pi/2$: it does not account for points in the second or third quadrants. This is why we drew a triangle and used its (positive) edge lengths to compute the angle φ:



$$\varphi = \tan^{-1}\left(\frac{1}{\sqrt{3}}\right) = \frac{\pi}{6}$$
$$\theta = \pi - \varphi = \frac{5\pi}{6}.$$

Figure 6.4.10

Alternatively, we could have observed that $\begin{pmatrix}-\sqrt{3}\\1\end{pmatrix}$ lies in the second quadrant, so that the angle θ in question is

$$\theta = \tan^{-1}\left(\frac{1}{-\sqrt{3}}\right) + \pi.$$

Note 6.4.4

When finding the rotation angle of a vector $\begin{pmatrix}a\\b\end{pmatrix}$, do not blindly compute $\tan^{-1}(b/a)$, since this will give the wrong answer when $\begin{pmatrix}a\\b\end{pmatrix}$ is in the second or third quadrant. Instead, draw a picture.

Geometry of 2×2 Matrices with a Complex Eigenvalue

Let A be a 2×2 matrix with a complex, non-real eigenvalue λ. Then A also has the eigenvalue $\bar{\lambda} \neq \lambda$. In particular, A has distinct eigenvalues, so it is diagonalizable using the complex numbers. We often like to think of our matrices as describing transformations of $\mathbb{R}^n$ (as opposed to $\mathbb{C}^n$). Because of this, the following construction is useful. It gives something like a diagonalization, except that all matrices involved have *real* entries.

Theorem 6.4.1: Rotation-Scaling Theorem

Let A be a 2×2 real matrix with a complex (non-real) eigenvalue λ, and let v be an eigenvector. Then $A = CBC^{-1}$ for

$$C = \begin{pmatrix} | & | \\ \Re(v) & \Im(v) \\ | & | \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} \Re(\lambda) & \Im(\lambda) \\ -\Im(\lambda) & \Re(\lambda) \end{pmatrix}.$$

In particular, A is similar to a rotation-scaling matrix that scales by a factor of $|\lambda| = \sqrt{\det(B)}$.

Proof

First we need to show that $\Re(v)$ and $\Im(v)$ are linearly independent, since otherwise C is not invertible. If not, then there exist real numbers x, y, not both equal to zero, such that $x\Re(v) + y\Im(v) = 0$. Then

$$\begin{aligned}(y+ix)v &= (y+ix)\big(\Re(v)+i\Im(v)\big)\\ &= y\Re(v) - x\Im(v) + \big(x\Re(v)+y\Im(v)\big)\,i\\ &= y\Re(v) - x\Im(v).\end{aligned}$$

Now, $(y+ix)v$ is also an eigenvector of A with eigenvalue λ, as it is a scalar multiple of v. But we just showed that $(y+ix)v$ is a vector with real entries, and any *real* eigenvector of a real matrix has a real eigenvalue. Therefore, $\Re(v)$ and $\Im(v)$ must be linearly independent after all.

Let $\lambda = a+bi$ and $v = \begin{pmatrix}x+yi\\z+wi\end{pmatrix}$. We observe that

$$\begin{aligned}Av = \lambda v &= (a+bi)\begin{pmatrix}x+yi\\z+wi\end{pmatrix}\\ &= \begin{pmatrix}(ax-by)+(ay+bx)i\\(az-bw)+(aw+bz)i\end{pmatrix}\\ &= \begin{pmatrix}ax-by\\az-bw\end{pmatrix} + i\begin{pmatrix}ay+bx\\aw+bz\end{pmatrix}.\end{aligned}$$

On the other hand, we have

$$A\left(\begin{pmatrix}x\\z\end{pmatrix}+i\begin{pmatrix}y\\w\end{pmatrix}\right)=A\begin{pmatrix}x\\z\end{pmatrix}+iA\begin{pmatrix}y\\w\end{pmatrix}=A\Re(v)+iA\Im(v).$$

Matching real and imaginary parts gives

$$A\Re(v)=\begin{pmatrix}ax-by\\az-bw\end{pmatrix}\qquad A\Im(v)=\begin{pmatrix}ay+bx\\aw+bz\end{pmatrix}.$$

Now we compute $CBC^{-1}\Re(v)$ and $CBC^{-1}\Im(v)$. Since $Ce_1=\Re(v)$ and $Ce_2=\Im(v)$, we have $C^{-1}\Re(v)=e_1$ and $C^{-1}\Im(v)=e_2$, so

$$\begin{aligned}CBC^{-1}\Re(v)&=CBe_1=C\begin{pmatrix}a\\-b\end{pmatrix}=a\Re(v)-b\Im(v)\\&=a\begin{pmatrix}x\\z\end{pmatrix}-b\begin{pmatrix}y\\w\end{pmatrix}=\begin{pmatrix}ax-by\\az-bw\end{pmatrix}=A\Re(v)\\CBC^{-1}\Im(v)&=CBe_2=C\begin{pmatrix}b\\a\end{pmatrix}=b\Re(v)+a\Im(v)\\&=b\begin{pmatrix}x\\z\end{pmatrix}+a\begin{pmatrix}y\\w\end{pmatrix}=\begin{pmatrix}ay+bx\\aw+bz\end{pmatrix}=A\Im(v).\end{aligned}$$

Therefore, $A\Re(v)=CBC^{-1}\Re(v)$ and $A\Im(v)=CBC^{-1}\Im(v)$.

Since $\Re(v)$ and $\Im(v)$ are linearly independent, they form a basis for $\mathbb{R}^2$. Let w be any vector in $\mathbb{R}^2$, and write $w=c\Re(v)+d\Im(v)$. Then

$$\begin{aligned}Aw&=A\big(c\Re(v)+d\Im(v)\big)\\&=cA\Re(v)+dA\Im(v)\\&=cCBC^{-1}\Re(v)+dCBC^{-1}\Im(v)\\&=CBC^{-1}\big(c\Re(v)+d\Im(v)\big)\\&=CBC^{-1}w.\end{aligned}$$

This proves that $A=CBC^{-1}$.

Here $\Re$ and $\Im$ denote the real and imaginary parts, respectively:

$$\Re(a+bi)=a\quad\Im(a+bi)=b\quad\Re\begin{pmatrix}x+yi\\z+wi\end{pmatrix}=\begin{pmatrix}x\\z\end{pmatrix}\quad\Im\begin{pmatrix}x+yi\\z+wi\end{pmatrix}=\begin{pmatrix}y\\w\end{pmatrix}.$$

The rotation-scaling matrix in question is the matrix

$$B=\begin{pmatrix}a&-b\\b&a\end{pmatrix}\quad\text{with}\quad a=\Re(\lambda),\ b=-\Im(\lambda).$$

Geometrically, the rotation-scaling theorem says that a 2×2 matrix with a complex eigenvalue behaves similarly to a rotation-scaling matrix. See Note 5.3.3 in Section 5.3.

One should regard Theorem 6.4.1 as a close analogue of Theorem 5.4.1 in Section 5.4, with a rotation-scaling matrix playing the role of a diagonal matrix. Before continuing, we restate the theorem as a recipe:

Recipe: A 2×2 Matrix with a Complex Eigenvalue

Let A be a 2×2 real matrix.

1. Compute the characteristic polynomial

$$f(\lambda)=\lambda^2-\operatorname{Tr}(A)\lambda+\det(A),$$

then compute its roots using the quadratic formula.

2. If the eigenvalues are complex, choose *one* of them, and call it λ.
3. Find a corresponding (complex) eigenvalue v using the trick 3.
4. Then $A = CBC^{-1}$ for

$$C = \begin{pmatrix} | & | \\ \Re(v) & \Im(v) \\ | & | \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} \Re(\lambda) & \Im(\lambda) \\ -\Im(\lambda) & \Re(\lambda) \end{pmatrix}.$$

This scales by a factor of $|\lambda|$.

Example 6.4.6

What does the matrix

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

do geometrically?

Solution

The eigenvalues of A are

$$\lambda = \frac{\operatorname{Tr}(A) \pm \sqrt{\operatorname{Tr}(A)^2 - 4\det(A)}}{2} = \frac{2 \pm \sqrt{4-8}}{2} = 1 \pm i.$$

We choose the eigenvalue $\lambda = 1 - i$ and find a corresponding eigenvector, using the trick, note 6.4.3:

$$A - (1-i)I_2 = \begin{pmatrix} 1+i & -1 \\ \star & \star \end{pmatrix} \xrightarrow{\text{eigenvector}} v = \begin{pmatrix} 1 \\ 1+i \end{pmatrix}.$$

According to Theorem 6.4.1, we have $A = CBC^{-1}$ for

$$C = \begin{pmatrix} \Re\begin{pmatrix} 1 \\ 1+i \end{pmatrix} & \Im\begin{pmatrix} 1 \\ 1+i \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$
$$B = \begin{pmatrix} \Re(\lambda) & \Im(\lambda) \\ -\Im(\lambda) & \Re(\lambda) \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

The matrix B is the rotation-scaling matrix in above Example 6.4.4: it rotates counterclockwise by an angle of 45° and scales by a factor of $\sqrt{2}$. The matrix A does the same thing, but with respect to the $\Re(v), \Im(v)$-coordinate system:

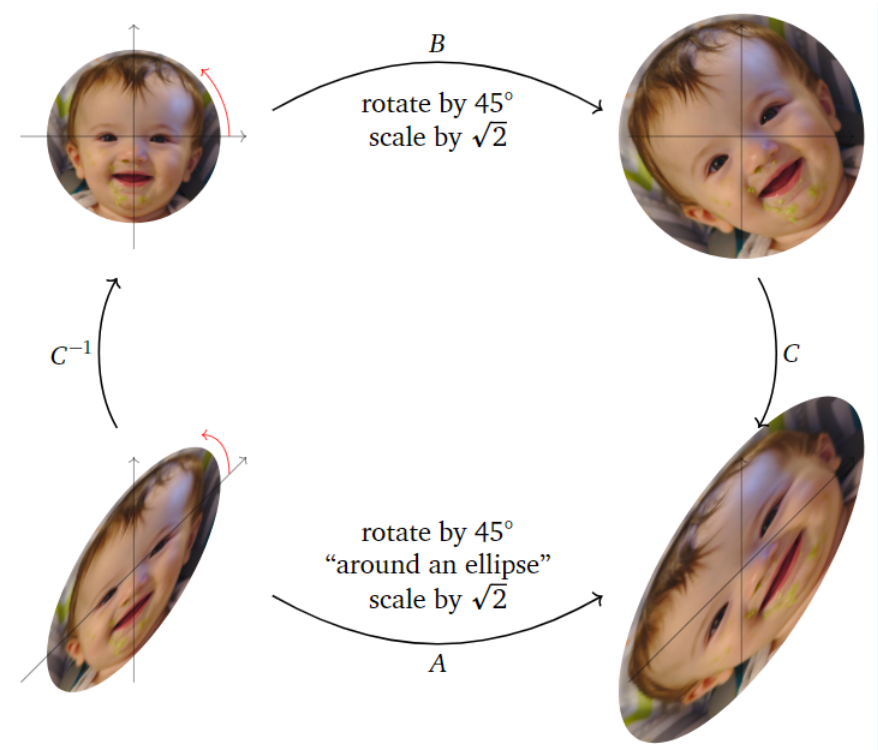


Figure 6.4.11

To summarize:

- B rotates around the circle centered at the origin and passing through e_1 and e_2, in the direction from e_1 to e_2, then scales by $\sqrt{2}$.
- A rotates around the ellipse centered at the origin and passing through $\Re(v)$ and $\Im(v)$, in the direction from $\Re(v)$ to $\Im(v)$, then scales by $\sqrt{2}$.

The reader might want to refer back to Example 5.3.7 in Section 5.3.

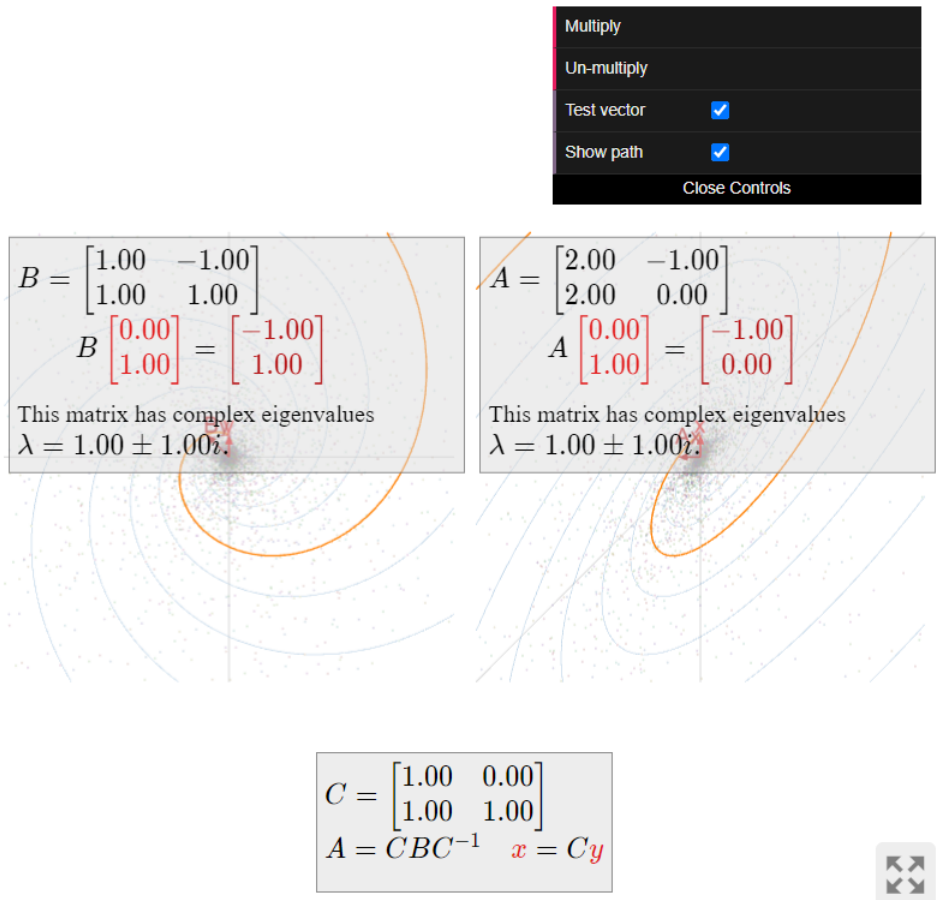


Figure 6.4.12: The geometric action of A and B on the plane. Click "multiply" to multiply the colored points by B on the left and A on the right.

If instead we had chosen $\bar{\lambda} = 1+i$ as our eigenvalue, then we would have found the eigenvector $\bar{v} = \binom{1}{1-i}$. In this case we would have $A = C'B'(C')^{-1}$, where

$$C' = \begin{pmatrix} \Re\begin{pmatrix} 1 \\ 1-i \end{pmatrix} & \Im\begin{pmatrix} 1 \\ 1-i \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}$$
$$B' = \begin{pmatrix} \Re(\bar{\lambda}) & \Im(\bar{\lambda}) \\ -\Im(\bar{\lambda}) & \Re(\bar{\lambda}) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

So, A is also similar to a *clockwise* rotation by 45°, followed by a scale by $\sqrt{2}$.

Example 6.4.7

What does the matrix

$$A = \begin{pmatrix} -\sqrt{3}+1 & -2 \\ 1 & -\sqrt{3}-1 \end{pmatrix}$$

do geometrically?

Solution

The eigenvalues of A are

$$\lambda = \frac{\operatorname{Tr}(A) \pm \sqrt{\operatorname{Tr}(A)^2 - 4\det(A)}}{2} = \frac{-2\sqrt{3} \pm \sqrt{12-16}}{2} = -\sqrt{3} \pm i.$$

We choose the eigenvalue $\lambda = -\sqrt{3} - i$ and find a corresponding eigenvector, using the trick, note 6.4.3:

$$A-(-\sqrt{3}-i)I_2=\begin{pmatrix}1+i & -2\\ \star & \star\end{pmatrix}\xrightarrow{\text{eigenvector}}v=\begin{pmatrix}2\\1+i\end{pmatrix}.$$

According to Theorem 6.4.1, we have $A=CBC^{-1}$ for

$$C=\begin{pmatrix}\Re\begin{pmatrix}2\\1+i\end{pmatrix} & \Im\begin{pmatrix}2\\1+i\end{pmatrix}\end{pmatrix}=\begin{pmatrix}2&0\\1&1\end{pmatrix}$$
$$B=\begin{pmatrix}\Re(\lambda) & \Im(\lambda)\\ -\Im(\lambda) & \Re(\lambda)\end{pmatrix}=\begin{pmatrix}-\sqrt{3} & -1\\ 1 & -\sqrt{3}\end{pmatrix}.$$

The matrix B is the rotation-scaling matrix in the above Example 6.4.5: it rotates counterclockwise by an angle of $5\pi/6$ and scales by a factor of 2. The matrix A does the same thing, but with respect to the $\Re(v), \Im(v)$-coordinate system:

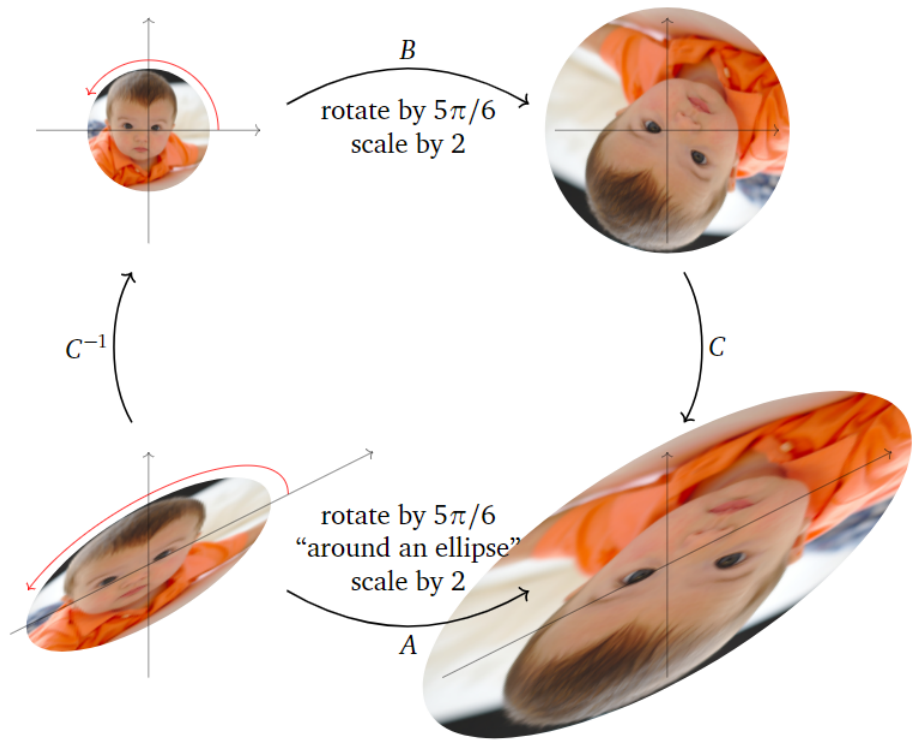


Figure 6.4.13

To summarize:

- B rotates around the circle centered at the origin and passing through e_1 and e_2, in the direction from e_1 to e_2, then scales by 2.
- A rotates around the ellipse centered at the origin and passing through $\Re(v)$ and $\Im(v)$, in the direction from $\Re(v)$ to $\Im(v)$, then scales by 2.

The reader might want to refer back to Example 5.3.7 in Section 5.3.

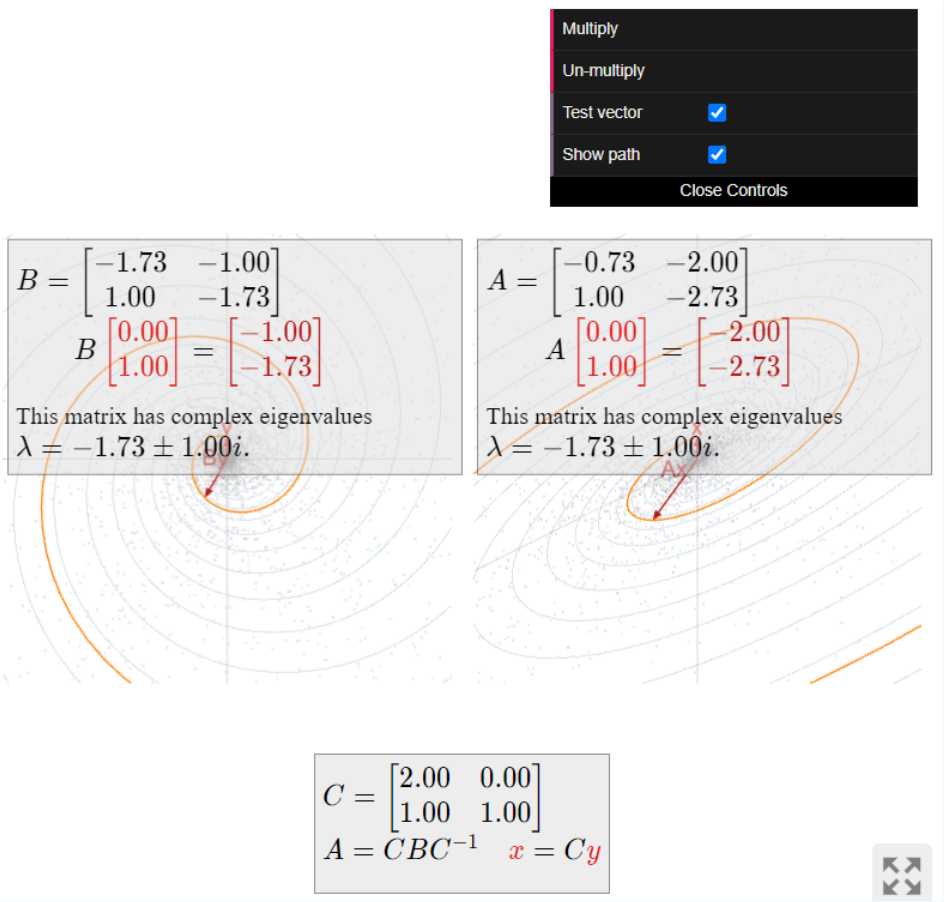


Figure 6.4.14: The geometric action of A and B on the plane. Click "multiply" to multiply the colored points by B on the left and A on the right.

If instead we had chosen $\bar{\lambda} = -\sqrt{3} - i$ as our eigenvalue, then we would have found the eigenvector $\bar{v} = \binom{2}{1-i}$. In this case we would have $A = C'B'(C')^{-1}$, where

$$C' = \begin{pmatrix} \Re\begin{pmatrix} 2 \\ 1-i \end{pmatrix} & \Im\begin{pmatrix} 2 \\ 1-i \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 1 & -1 \end{pmatrix}$$
$$B' = \begin{pmatrix} \Re(\bar{\lambda}) & \Im(\bar{\lambda}) \\ -\Im(\bar{\lambda}) & \Re(\bar{\lambda}) \end{pmatrix} = \begin{pmatrix} -\sqrt{3} & 1 \\ -1 & -\sqrt{3} \end{pmatrix}.$$

So, A is also similar to a *clockwise* rotation by $5\pi/6$, followed by a scale by 2.

We saw in the above examples that Theorem 6.4.1 can be applied in two different ways to any given matrix: one has to choose one of the two conjugate eigenvalues to work with. Replacing λ by $\bar{\lambda}$ has the effect of replacing v by $\bar{v}$, which just negates all imaginary parts, so we also have $A = C'B'(C')^{-1}$ for

$$C' = \begin{pmatrix} | & | \\ \Re(v) & -\Im(v) \\ | & | \end{pmatrix} \quad \text{and} \quad B' = \begin{pmatrix} \Re(\lambda) & -\Im(\lambda) \\ \Im(\lambda) & \Re(\lambda) \end{pmatrix}.$$

The matrices B and B' are similar to each other. The only difference between them is the direction of rotation, since $\binom{\Re(\lambda)}{-\Im(\lambda)}$ and $\binom{\Re(\lambda)}{\Im(\lambda)}$ are mirror images of each other over the x-axis:

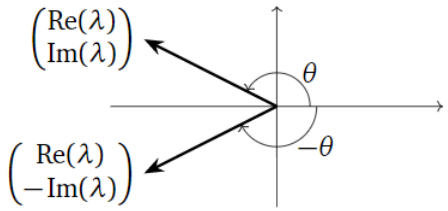


Figure 6.4.15

The discussion that follows is closely analogous to the exposition in subsection The Geometry of Diagonalizable Matrices in Section 5.4, in which we studied the dynamics of diagonalizable 2×2 matrices.

Note 6.4.5: Dynamics of a 2×2 Matrix with a Complex Eigenvalue

Let A be a 2×2 matrix with a complex (non-real) eigenvalue λ. By Theorem 6.4.1, the matrix A is similar to a matrix that rotates by some amount and scales by $|\lambda|$. Hence, A rotates around an ellipse and scales by $|\lambda|$. There are three different cases.

$|\lambda| > 1$: when the scaling factor is greater than 1, then vectors tend to get longer, i.e., farther from the origin. In this case, repeatedly multiplying a vector by A makes the vector "spiral out". For example,

$$A = \frac{1}{\sqrt{2}}\begin{pmatrix}\sqrt{3}+1 & -2\\ 1 & \sqrt{3}-1\end{pmatrix} \qquad \lambda = \frac{\sqrt{3}-i}{\sqrt{2}} \qquad |\lambda| = \sqrt{2} > 1$$

gives rise to the following picture:

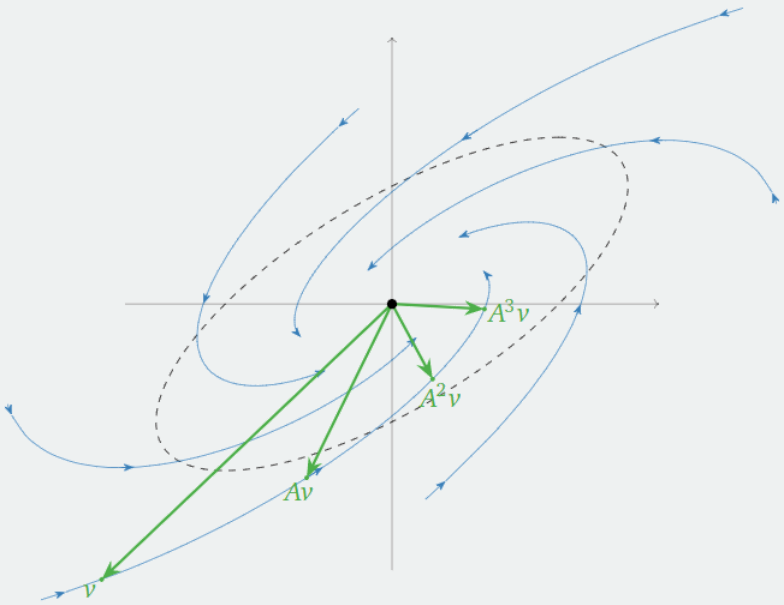


Figure 6.4.16

$|\lambda| = 1$: when the scaling factor is equal to 1, then vectors do not tend to get longer or shorter. In this case, repeatedly multiplying a vector by A simply "rotates around an ellipse". For example,

$$A = \frac{1}{2}\begin{pmatrix}\sqrt{3}+1 & -2\\ 1 & \sqrt{3}-1\end{pmatrix} \qquad \lambda = \frac{\sqrt{3}-i}{2} \qquad |\lambda| = 1$$

gives rise to the following picture:

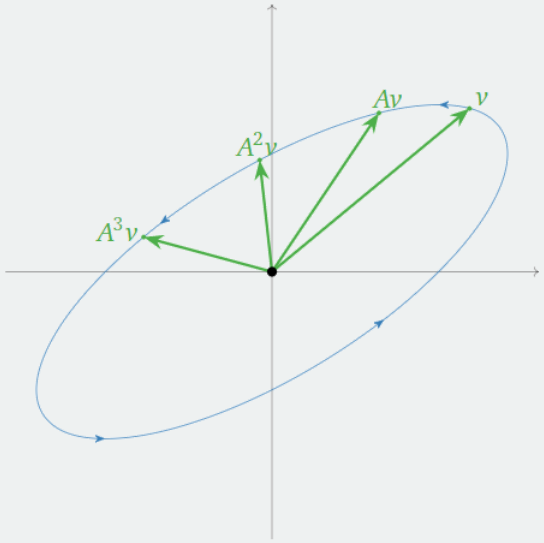


Figure 6.4.17

$|\lambda| < 1$: when the scaling factor is less than 1, then vectors tend to get shorter, i.e., closer to the origin. In this case, repeatedly multiplying a vector by A makes the vector "spiral in". For example,

$$A = \frac{1}{2\sqrt{2}}\begin{pmatrix} \sqrt{3}+1 & -2 \\ 1 & \sqrt{3}-1 \end{pmatrix} \qquad \lambda = \frac{\sqrt{3}-i}{2\sqrt{2}} \qquad |\lambda| = \frac{1}{\sqrt{2}} < 1$$

gives rise to the following picture:

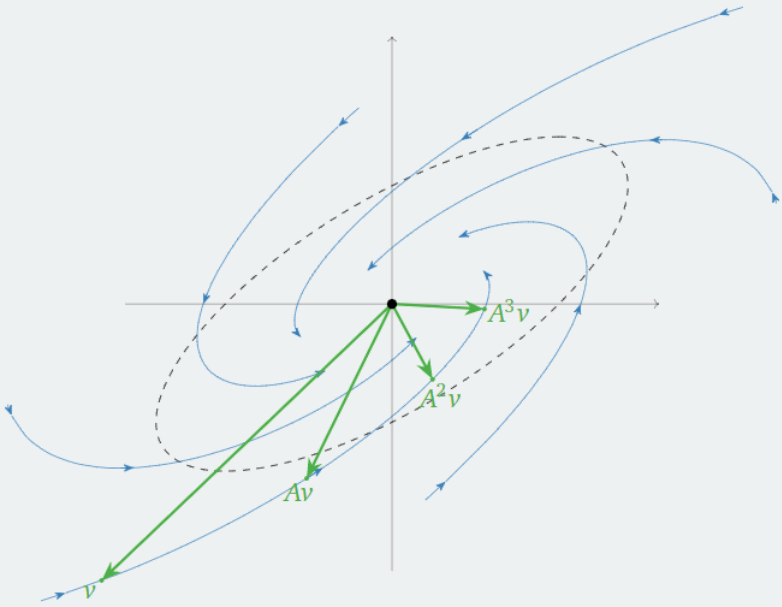


Figure 6.4.18

Example 6.4.8: Interactive: $|\lambda| > 1$

$$A = \frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{3}+1 & -2 \\ 1 & \sqrt{3}-1 \end{pmatrix} \qquad B = \frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{pmatrix} \qquad C = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$$

$$\lambda = \frac{\sqrt{3}-i}{\sqrt{2}} \qquad |\lambda| = \sqrt{2} > 1$$

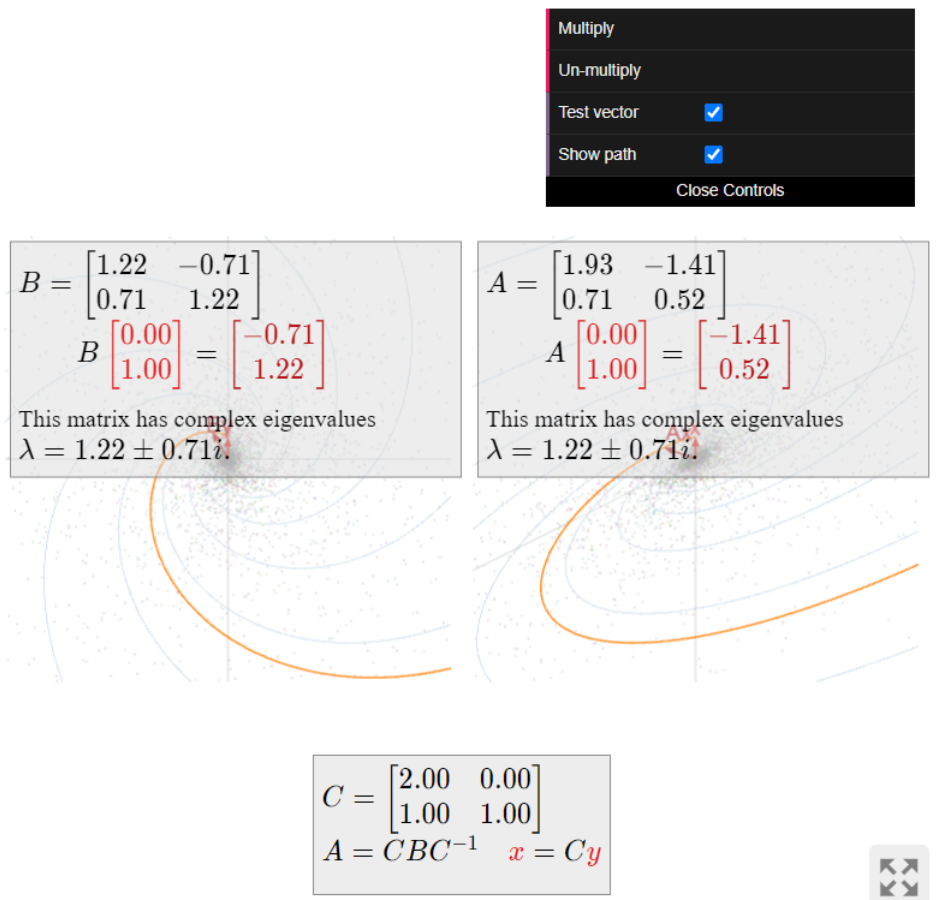


Figure 6.4.19: The geometric action of A and B on the plane. Click “multiply” to multiply the colored points by B on the left and A on the right.

Example 6.4.9: Interactive: $|\lambda| = 1$

$$A = \frac{1}{2}\begin{pmatrix} \sqrt{3}+1 & -2 \\ 1 & \sqrt{3}-1 \end{pmatrix} \qquad B = \frac{1}{2}\begin{pmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{pmatrix} \qquad C = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$$

$$\lambda = \frac{\sqrt{3}-i}{2} \qquad |\lambda| = 1$$

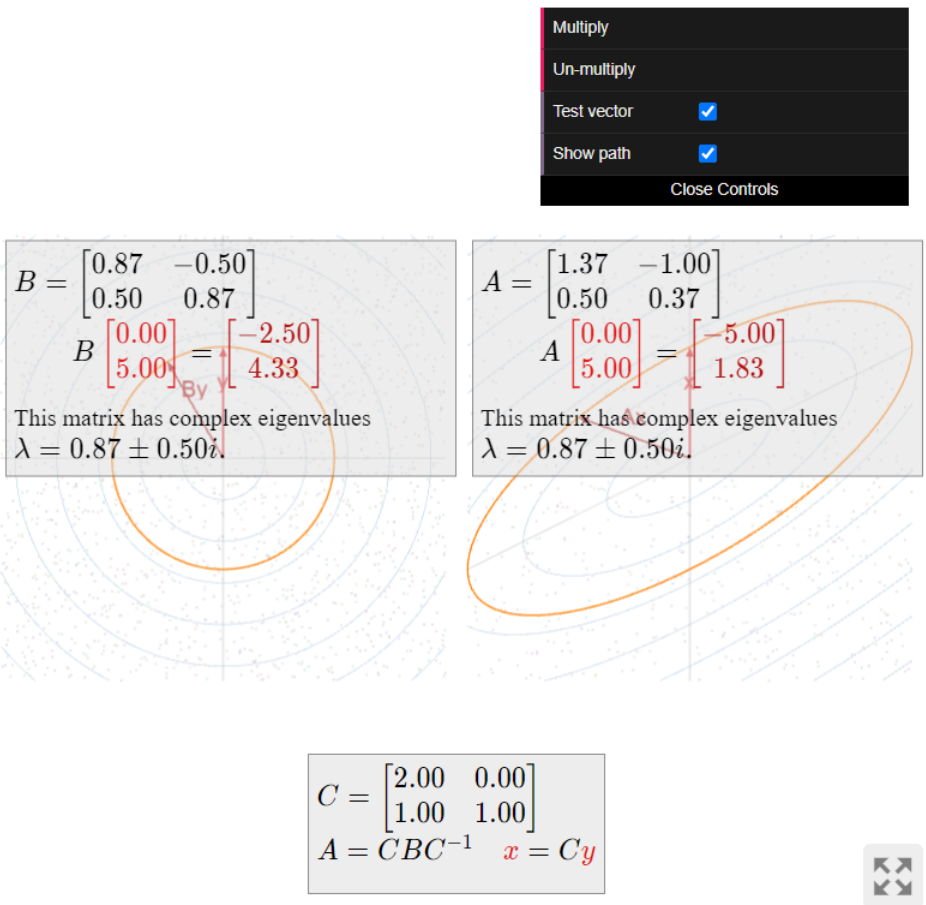


Figure 6.4.20: The geometric action of A and B on the plane. Click “multiply” to multiply the colored points by B on the left and A on the right.

Example 6.4.10: Interactive: $|\lambda| < 1$

$$A = \frac{1}{2\sqrt{2}} \begin{pmatrix} \sqrt{3}+1 & -2 \\ 1 & \sqrt{3}-1 \end{pmatrix} \qquad B = \frac{1}{2\sqrt{2}} \begin{pmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{pmatrix} \qquad C = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$$

$$\lambda = \frac{\sqrt{3}-i}{2\sqrt{2}} \qquad |\lambda| = \frac{1}{\sqrt{2}} < 1$$

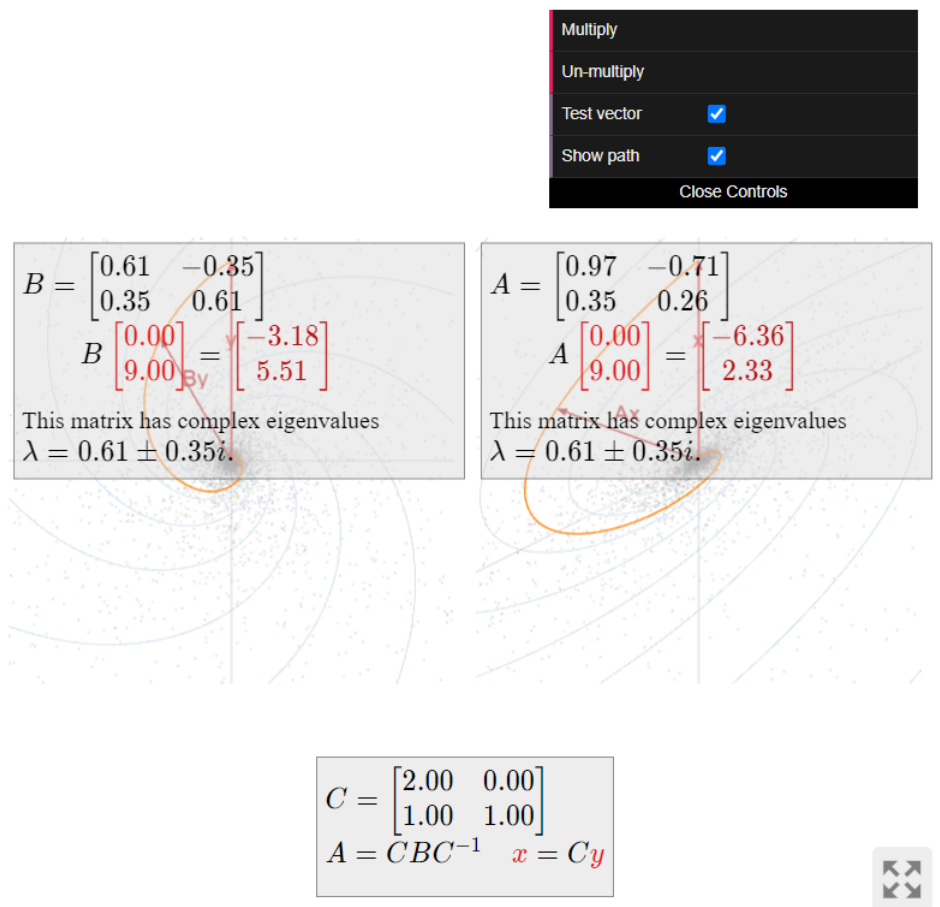


Figure 6.4.21: The geometric action of A and B on the plane. Click "multiply" to multiply the colored points by B on the left and A on the right.

Remark: Classification of 2×2 matrices up to similarity

At this point, we can write down the "simplest" possible matrix which is similar to any given 2×2 matrix A. There are four cases:

1. A has two real eigenvalues λ_1, λ_2. In this case, A is diagonalizable, so A is similar to the matrix
$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$
This representation is unique up to reordering the eigenvalues.
2. A has one real eigenvalue λ of geometric multiplicity 2. In this case, we saw in Example 5.4.20 in Section 5.4 that A is *equal* to the matrix
$$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}.$$
3. A has one real eigenvalue λ of geometric multiplicity 1. In this case, A is not diagonalizable, and we saw in Remark: Non-diagonalizable 2×2 matrices with an eigenvalue in Section 5.4 that A is similar to the matrix
$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}.$$
4. A has no real eigenvalues. In this case, A has a complex eigenvalue λ, and A is similar to the rotation-scaling matrix

$$\begin{pmatrix} \Re(\lambda) & \Im(\lambda) \\ -\Im(\lambda) & \Re(\lambda) \end{pmatrix}$$

by Theorem 6.4.1. By Proposition 6.4.1, the eigenvalues of a rotation-scaling matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ $are\backslash(a \pm bi$, so that two rotation-scaling matrices $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ and $\begin{pmatrix} c & -d \\ d & c \end{pmatrix}$ are similar if and only if $a = c$ and $b = \pm d$.

Block Diagonalization

For matrices larger than 2×2, there is a theorem that combines Theorem 5.4.1 in Section 5.4 and Theorem 6.4.1. It says essentially that a matrix is similar to a matrix with parts that look like a diagonal matrix, and parts that look like a rotation-scaling matrix.

Theorem 6.4.2: Block Diagonalization Theorem

Let A be a real $n \times n$ matrix. Suppose that for each (real or complex) eigenvalue, the algebraic multiplicity equals the geometric multiplicity. Then $A = CBC^{-1}$, where B and C are as follows:

- The matrix B is ***block diagonal***, where the blocks are 1×1 blocks containing the real eigenvalues (with their multiplicities), or 2×2 blocks containing the matrices

$$\begin{pmatrix} \Re(\lambda) & \Im(\lambda) \\ -\Im(\lambda) & \Re(\lambda) \end{pmatrix}$$

  for each non-real eigenvalue λ (with multiplicity).
- The columns of C form bases for the eigenspaces for the real eigenvectors, or come in pairs $\begin{pmatrix} \Re(v) & \Im(v) \end{pmatrix}$ for the non-real eigenvectors.

The Theorem 6.4.2 is proved in the same way as Theorem 5.4.1 in Section 5.4 and Theorem 6.4.1. It is best understood in the case of 3×3 matrices.

Note 6.4.6: Block Diagonalization of a 3×3 Matrix with a Complex Eigenvalue

Let A be a 3×3 matrix with a complex eigenvalue λ_1. Then $\bar{\lambda}_1$ is another eigenvalue, and there is one real eigenvalue λ_2. Since there are three distinct eigenvalues, they have algebraic and geometric multiplicity one, so Theorem 6.4.2 applies to A.

Let v_1 be a (complex) eigenvector with eigenvalue λ_1, and let v_2 be a (real) eigenvector with eigenvalue λ_2. Then Theorem 6.4.2 says that $A = CBC^{-1}$ for

$$C = \begin{pmatrix} | & | & | \\ \mathrm{Re}(v_1) & \mathrm{Im}(v_1) & v_2 \\ | & | & | \end{pmatrix} \qquad B = \begin{pmatrix} \mathrm{Re}(\lambda_1) & \mathrm{Im}(\lambda_1) & 0 \\ -\mathrm{Im}(\lambda_1) & \mathrm{Re}(\lambda_1) & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix}.$$

Figure 6.4.22

Example 6.4.11: Geometry of a 3×3 matrix with a complex eigenvalue

What does the matrix

$$A = \frac{1}{29}\begin{pmatrix} 33 & -23 & 9 \\ 22 & 33 & -23 \\ 19 & 14 & 50 \end{pmatrix}$$

do geometrically?

Solution

First we find the (real and complex) eigenvalues of A. We compute the characteristic polynomial using whatever method we like:

$$f(\lambda) = \det(A - \lambda I_3) = -\lambda^3 + 4\lambda^2 - 6\lambda + 4.$$

We search for a real root using the rational root theorem. The possible rational roots are $\pm 1, \pm 2, \pm 4$; we find $f(2) = 0$, so that $\lambda - 2$ divides $f(\lambda)$. Performing polynomial long division gives

$$f(\lambda) = -(\lambda - 2)\big(\lambda^2 - 2\lambda + 2\big).$$

The quadratic term has roots

$$\lambda = \frac{2 \pm \sqrt{4-8}}{2} = 1 \pm i,$$

so that the complete list of eigenvalues is $\lambda_1 = 1 - i,\ \bar{\lambda}_1 = 1 + i,$ and $\lambda_2 = 2$.

Now we compute some eigenvectors, starting with $\lambda_1 = 1 - i$. We row reduce (probably with the aid of a computer):

$$A - (1-i)I_3 = \frac{1}{29}\begin{pmatrix} 4+29i & -23 & 9 \\ 22 & 4+29i & -23 \\ 19 & 14 & 21+29i \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 7/5 + i/5 \\ 0 & 1 & -2/5 + 9i/5 \\ 0 & 0 & 0 \end{pmatrix}.$$

The free variable is z, and the parametric form is

$$\begin{cases} x &= -\left(\frac{7}{5} + \frac{1}{5}i\right) z \\ y &= \left(\frac{2}{5} - \frac{9}{5}i\right) z \end{cases} \xrightarrow[\text{eigenvector}]{z=5} v_1 = \begin{pmatrix} -7 - i \\ 2 - 9i \\ 5 \end{pmatrix}.$$

For $\lambda_2 = 2,$ we have

$$A - 2I_3 = \frac{1}{29}\begin{pmatrix} -25 & -23 & 9 \\ 22 & -25 & -23 \\ 19 & 14 & -8 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & -2/3 \\ 0 & 1 & 1/3 \\ 0 & 0 & 0 \end{pmatrix}.$$

The free variable is z, and the parametric form is

$$\begin{cases} x &= \frac{2}{3} z \\ y &= -\frac{1}{3} z \end{cases} \xrightarrow[\text{eigenvector}]{z=3} v_2 = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}.$$

According to Theorem 6.4.2, we have $A = CBC^{-1}$ for

$$C = \begin{pmatrix} | & | & | \\ \Re(v_1) & \Im(v_1) & v_2 \\ | & | & | \end{pmatrix} = \begin{pmatrix} -7 & -1 & 2 \\ 2 & -9 & -1 \\ 5 & 0 & 3 \end{pmatrix}$$

$$B = \begin{pmatrix} \Re(\lambda_1) & \Im(\lambda_1) & 0 \\ -\Im(\lambda_1) & \Re(\lambda_1) & 0 \\ 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

The matrix B is a combination of the rotation-scaling matrix $\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ from Example 6.4.4, and a diagonal matrix. More specifically, B acts on the xy-coordinates by rotating counterclockwise by 45° and scaling by $\sqrt{2}$, and it scales the z-coordinate by 2. This means that points above the xy-plane spiral out away from the z-axis and move up, and points below the xy-plane spiral out away from the z-axis and move down.

The matrix A does the same thing as B, but with respect to the $\{\Re(v_1), \Im(v_1), v_2\}$-coordinate system. That is, A acts on the $\Re(v_1), \Im(v_1)$-plane by spiraling out, and A acts on the v_2-coordinate by scaling by a factor of 2. See the demo below.

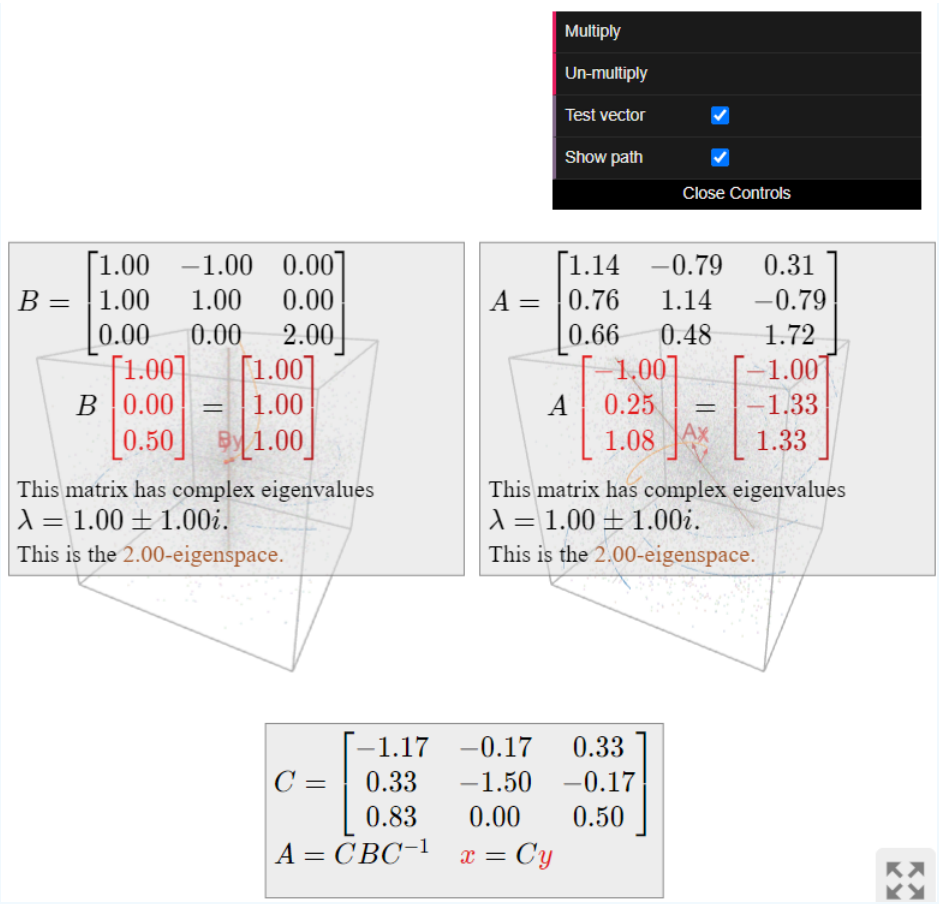


Figure 6.4.23: The geometric action of A and B on $\mathbb{R}^3$. Click "multiply" to multiply the colored points by B on the left and A on the right. (We have scaled C by $1/6$ so that the vectors x and y have roughly the same size.)

This page titled 6.4: Complex Eigenvalues is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **5.5: Complex Eigenvalues** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

6.5: Stochastic Matrices

Objectives

1. Learn examples of stochastic matrices and applications to difference equations.
2. Understand Google's PageRank algorithm.
3. *Recipe:* find the steady state of a positive stochastic matrix.
4. *Picture:* dynamics of a positive stochastic matrix.
5. *Theorem:* the Perron–Frobenius theorem.
6. *Vocabulary words:* ***difference equation, (positive) stochastic matrix, steady state, importance matrix, Google matrix***.

This section is devoted to one common kind of application of eigenvalues: to the study of difference equations, in particular to Markov chains. We will introduce stochastic matrices, which encode this type of difference equation, and will cover in detail the most famous example of a stochastic matrix: the Google Matrix.

Difference Equations

Suppose that we are studying a system whose state at any given time can be described by a list of numbers: for instance, the numbers of rabbits aged $0, 1$, and 2 years, respectively, or the number of copies of Prognosis Negative in each of the Red Box kiosks in Atlanta. In each case, we can represent the state at time t by a vector v_t. We assume that t represents a discrete time quantity: in other words, v_t is the state "on day t" or "at year t". Suppose in addition that the state at time $t+1$ is related to the state at time t in a linear way: $v_{t+1} = Av_t$ for some matrix A. This is the situation we will consider in this subsection.

Definition 6.5.1: Difference Equation

A ***difference equation*** is an equation of the form

$$v_{t+1} = Av_t$$

for an $n \times n$ matrix A and vectors $v_0, v_1, v_2, \ldots$ in $\mathbb{R}^n$.

In other words:

- v_t is the "state at time t,"
- v_{t+1} is the "state at time $t+1$," and
- $v_{t+1} = Av_t$ means that A is the "change of state matrix."

Note that

$$v_t = Av_{t-1} = A^2v_{t-2} = \cdots = A^tv_0,$$

which should hint to you that the long-term behavior of a difference equation is an eigenvalue problem.

We will use the following example in this subsection and the next. Understanding this section amounts to understanding this example.

Example 6.5.1

Red Box has kiosks all over Atlanta where you can rent movies. You can return them to any other kiosk. For simplicity, pretend that there are three kiosks in Atlanta, and that every customer returns their movie the next day. Let v_t be the vector whose entries x_t, y_t, z_t are the number of copies of Prognosis Negative at kiosks $1, 2$, and 3, respectively. Let A be the matrix whose i, j-entry is the probability that a customer renting Prognosis Negative from kiosk j returns it to kiosk i. For example, the matrix

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

encodes a 30% probability that a customer renting from kiosk 3 returns the movie to kiosk 2, and a 40% probability that a movie rented from kiosk 1 gets returned to kiosk 3. The second row (for instance) of the matrix A says:

> *The number of movies returned to kiosk 2 will be (on average):*
>
> $$30\% \text{ of the movies from kiosk } 1$$
> $$40\% \text{ of the movies from kiosk } 2$$
> $$30\% \text{ of the movies from kiosk } 3$$

Applying this to all three rows, this means

$$A\begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} = \begin{pmatrix} .3x_t + .4y_t + .5z_t \\ .3x_t + .4y_t + .3z_t \\ .4x_t + .2y_t + .2z_t \end{pmatrix}.$$

Therefore, Av_t represents the number of movies in each kiosk the next day:

$$Av_t = v_{t+1}.$$

This system is modeled by a difference equation.

An important question to ask about a difference equation is: what is its long-term behavior? How many movies will be in each kiosk after 100 days? In the next subsection, we will answer this question for a particular type of difference equation.

Example 6.5.2: Rabbit population

In a population of rabbits,

1. half of the newborn rabbits survive their first year;
2. of those, half survive their second year;
3. the maximum life span is three years;
4. rabbits produce 0, 6, 8 rabbits in their first, second, and third years, respectively.

Let v_t be the vector whose entries x_t, y_t, z_t are the number of rabbits aged $0, 1$, and 2, respectively. The rules above can be written as a system of equations:

$$\begin{aligned} x_{t+1} &= 6y_t + 8z_t \\ y_{t+1} &= \tfrac{1}{2}x_t \\ z_{t+1} &= \tfrac{1}{2}y_t. \end{aligned}$$

In matrix form, this says:

$$\begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix} v_t = v_{t+1}.$$

This system is modeled by a difference equation.

Define

$$A = \begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

We compute A has eigenvalues 2 and -1, and that an eigenvector with eigenvalue 2 is

$$v = \begin{pmatrix} 16 \\ 4 \\ 1 \end{pmatrix}.$$

This partially explains why the ratio $x_t : y_t : z_t$ approaches $16 : 4 : 1$ and why all three quantities eventually double each year in this demo:

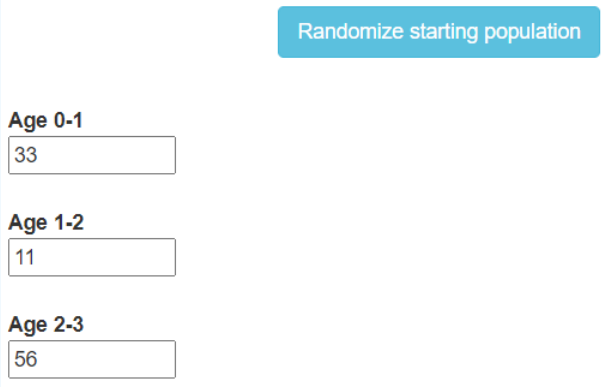


Figure 6.5.1: Left: the population of rabbits in a given year. Right: the proportions of rabbits in that year. Choose any values you like for the starting population, and click "Advance 1 year" several times. Notice that the ratio $x_t : y_t : z_t$ approaches $16 : 4 : 1$, and that all three quantities eventually double each year.

Stochastic Matrices and the Steady State

In this subsection, we discuss difference equations representing *probabilities*, like the Red Box example 6.5.1. Such systems are called *Markov chains*. The most important result in this section is the Perron–Frobenius theorem, which describes the long-term behavior of a Markov chain.

Definition 6.5.2: Stochastic

A square matrix A is ***stochastic*** if all of its entries are nonnegative, and the entries of each column sum to 1.

A matrix is *positive* if all of its entries are positive numbers.

Note 6.5.1

A positive stochastic matrix is a stochastic matrix whose entries are all positive numbers. In particular, no entry is equal to zero. For instance, the first matrix below is a positive stochastic matrix, and the second is not:

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

Remark

More generally, a ***regular*** stochastic matrix is a stochastic matrix A such that A^n is positive for some $n \geq 1$. The Perron–Frobenius theorem below also applies to regular stochastic matrices.

Example 6.5.3

Continuing with Red Box example 6.5.1, the matrix

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

is a positive stochastic matrix. The fact that the columns sum to 1 says that all of the movies rented from a particular kiosk must be returned to *some* other kiosk (remember that every customer returns their movie the next day). For instance, the first column says:

> *Of the movies rented from kiosk* 1,
>
> $$30\% \text{ will be returned to kiosk } 1$$
> $$30\% \text{ will be returned to kiosk } 2$$
> $$40\% \text{ will be returned to kiosk } 3.$$
>
> *The sum is* 100%*, as all of the movies are returned to one of the three kiosks.*

The matrix A represents the change of state from one day to the next:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \\ z_{t+1} \end{pmatrix} = A \begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} = \begin{pmatrix} .3x_t + .4y_t + .5z_t \\ .3x_t + .4y_t + .3z_t \\ .4x_t + .2y_t + .2z_t \end{pmatrix}.$$

If we sum the entries of v_{t+1}, we obtain

$$\begin{aligned} &(.3x_t + .4y_t + .5z_t) + (.3x_t + .4y_t + .3z_t) + (.4x_t + .2y_t + .2z_t) \\ &\quad = (.3 + .3 + .4)x_t + (.4 + .4 + .2)y_t + (.5 + .3 + .2)z_t \\ &\quad = x_t + y_t + z_t. \end{aligned}$$

This says that the *total* number of copies of Prognosis Negative in the three kiosks does not change from day to day, as we expect.

The fact that the entries of the vectors v_t and v_{t+1} sum to the same number is a consequence of the fact that the columns of a stochastic matrix sum to 1.

Note 6.5.2

Let A be a stochastic matrix, let v_t be a vector, and let $v_{t+1} = Av_t$. Then the sum of the entries of v_t equals the sum of the entries of v_{t+1}.

Computing the long-term behavior of a difference equation turns out to be an eigenvalue problem. The eigenvalues of stochastic matrices have very special properties.

Fact 6.5.1

Let A be a stochastic matrix. Then:

1. 1 is an eigenvalue of A.
2. If λ is a (real or complex) eigenvalue of A, then $|\lambda| \leq 1$.

Proof

If A is stochastic, then the *rows* of A^T sum to 1. But multiplying a matrix by the vector $(1, 1, \ldots, 1)$ sums the rows:

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

Therefore, 1 is an eigenvalue of A^T. But A and A^T have the same characteristic polynomial:

$$\det(A - \lambda I_n) = \det\big((A - \lambda I_n)^T\big) = \det(A^T - \lambda I_n).$$

Therefore, 1 is an eigenvalue of A.

Now let λ be any eigenvalue of A, so it is also an eigenvalue of A^T. Let $x = (x_1, x_2, \ldots, x_n)$ be an eigenvector of A^T with eigenvalue λ, so $\lambda x = A^T x$. The jth entry of this vector equation is

$$\lambda x_j = \sum_{i=1}^{n} a_{ij} x_i.$$

Choose x_j with the largest absolute value, so $|x_i| \leq |x_j|$ for all i. Then

$$|\lambda| \cdot |x_j| = \left| \sum_{i=1}^{n} a_{ij} x_i \right| \leq \sum_{i=1}^{n} a_{ij} \cdot |x_i| \leq \sum_{i=1}^{n} a_{ij} \cdot |x_j| = 1 \cdot |x_j|,$$

where the last equality holds because $\sum_{i=1}^{n} a_{ij} = 1$. This implies $|\lambda| \leq 1$.

In fact, for a *positive* stochastic matrix A, one can show that if $\lambda \neq 1$ is a (real or complex) eigenvalue of A, then $|\lambda| < 1$. The 1-eigenspace of a stochastic matrix is very important.

Definition 6.5.3: Steady State

A *steady state* of a stochastic matrix A is an eigenvector w with eigenvalue 1, such that the entries are *positive* and sum to 1.

The Perron–Frobenius theorem describes the long-term behavior of a difference equation represented by a stochastic matrix. Its proof is beyond the scope of this text.

Theorem 6.5.1: Perron-Frobenius Theorem

Let A be a positive stochastic matrix. Then A admits a unique steady state vector w, which spans the 1-eigenspace.

Moreover, for any vector v_0 with entries summing to some number c, the iterates

$$v_1 = Av_0, \ v_2 = Av_1, \ \ldots, \ v_t = Av_{t-1}, \ \ldots$$

approach cw as t gets large.

Translation: The Perron–Frobenius theorem makes the following assertions:

- The 1-eigenspace of a positive stochastic matrix is a line.
- The 1-eigenspace contains a vector with positive entries.
- All vectors approach the 1-eigenspace upon repeated multiplication by A.

One should think of a steady state vector w as a vector of *percentages*. For example, if the movies are distributed according to these percentages today, then they will be have the same distribution tomorrow, since $Aw = w$. And no matter the starting distribution of movies, the long-term distribution will always be the steady state vector.

The sum c of the entries of v_0 is the *total number* of things in the system being modeled. The total number does not change, so the long-term state of the system must approach cw: it is a multiple of w because it is contained in the 1-eigenspace, and the entries of cw sum to c.

Recipe 1: Compute the Steady State Vector

Let A be a positive stochastic matrix. Here is how to compute the steady-state vector of A.

1. Find any eigenvector v of A with eigenvalue 1 by solving $(A - I_n)v = 0$.
2. Divide v by the sum of the entries of v to obtain a vector w whose entries sum to 1.
3. This vector automatically has positive entries. It is the unique steady-state vector.

The above recipe is suitable for calculations by hand, but it does not take advantage of the fact that A is a stochastic matrix. In practice, it is generally faster to compute a steady state vector by computer as follows:

Recipe 2: Approximate the Steady State Vector by Computer

Let A be a positive stochastic matrix. Here is how to approximate the steady-state vector of A with a computer.

1. Choose any vector v_0 whose entries sum to 1 (e.g., a standard coordinate vector).
2. Compute $v_1 = Av_0$, $v_2 = Av_1$, $v_3 = Av_2$, etc.
3. These converge to the steady state vector w.

Example 6.5.4: A 2×2 stochastic matrix

Consider the positive stochastic matrix

$$A = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}.$$

This matrix has characteristic polynomial

$$f(\lambda) = \lambda^2 - \operatorname{Tr}(A)\lambda + \det(\lambda) = \lambda^2 - \frac{3}{2}\lambda + \frac{1}{2} = (\lambda - 1)(\lambda - 1/2).$$

Notice that 1 is strictly greater than the other eigenvalue, and that it has algebraic (hence, geometric) multiplicity 1. We compute eigenvectors for the eigenvalues $1, 1/2$ to be, respectively,

$$u_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad u_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

The eigenvector u_1 necessarily has positive entries; the steady-state vector is

$$w = \frac{1}{1+1}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 50\% \\ 50\% \end{pmatrix}.$$

The Perron–Frobenius theorem asserts that, for any vector v_0, the vectors $v_1 = Av_0$, $v_2 = Av_1$, $\ldots$ approach a vector whose entries are the same: 50% of the sum will be in the first entry, and 50% will be in the second.

We can see this explicitly, as follows. The eigenvectors u_1, u_2 form a basis $\mathcal{B}$ for $\mathbb{R}^2$; for any vector $x = a_1u_1 + a_2u_2$ in $\mathbb{R}^2$, we have

$$Ax = A(a_1u_1 + a_2u_2) = a_1Au_1 + a_2Au_2 = a_1u_1 + \frac{a_2}{2}u_2.$$

Iterating multiplication by A in this way, we have

$$A^t x = a_1u_1 + \frac{a_2}{2^t}u_2 \;\to\; a_1u_1$$

as $t \to \infty$. This shows that $A^t x$ approaches

$$a_1u_1 = \begin{pmatrix} a_1 \\ a_1 \end{pmatrix}.$$

Note that the sum of the entries of a_1u_1 is equal to the sum of the entries of $a_1u_1 + a_2u_2$, since the entries of u_2 sum to 0.

To illustrate the theorem with numbers, let us choose a particular value for u_0, say $u_0 = \binom{1}{0}$. We compute the values for $u_t = \binom{x_t}{y_t}$ in this table:

t	x_t	y_t
0	1.000	0.000
1	0.750	0.250
2	0.625	0.375
3	0.563	0.438
4	0.531	0.469
5	0.516	0.484
6	0.508	0.492
7	0.504	0.496
8	0.502	0.498
9	0.501	0.499
10	0.500	0.500

We see that u_t does indeed approach $\binom{0.5}{0.5}$.

Now we turn to visualizing the dynamics of (i.e., repeated multiplication by) the matrix A. This matrix is diagonalizable; we have $A = CDC^{-1}$ for

$$C = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

The matrix D leaves the x-coordinate unchanged and scales the y-coordinate by $1/2$. Repeated multiplication by D makes the y-coordinate very small, so it "sucks all vectors into the x-axis."

The matrix A does the same thing as D, but with respect to the coordinate system defined by the columns u_1, u_2 of C. This means that A "sucks all vectors into the 1-eigenspace", without changing the sum of the entries of the vectors.

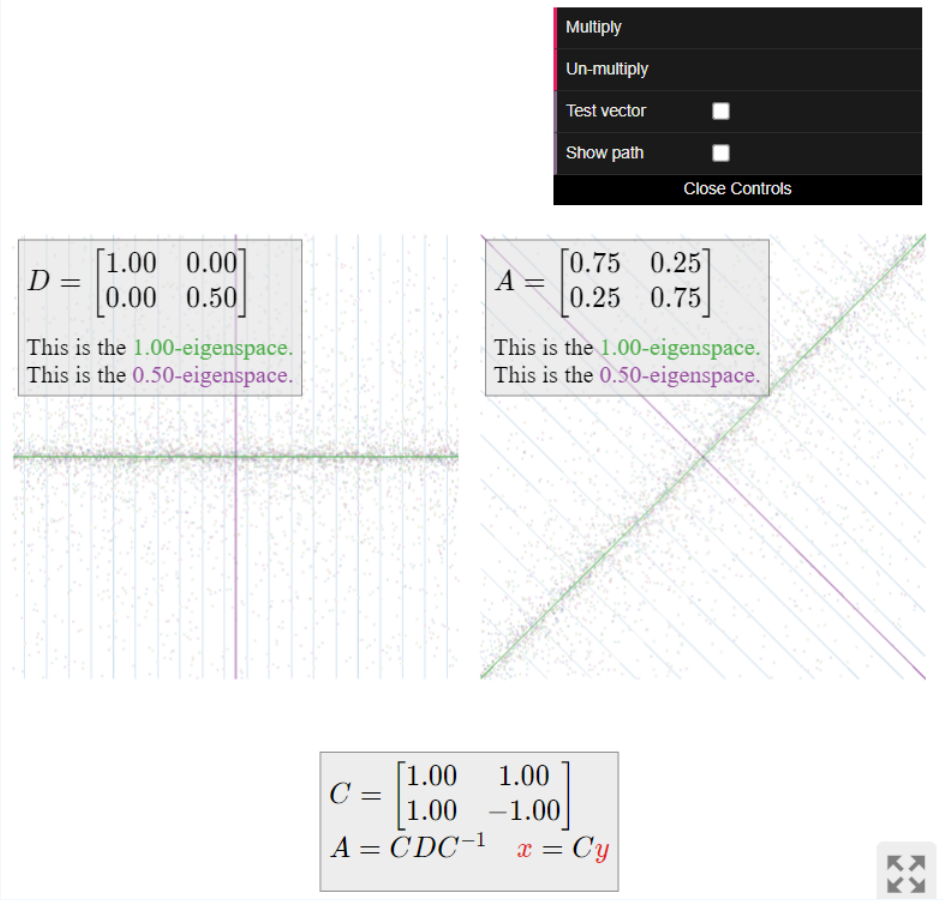


Figure 6.5.2: Dynamics of the stochastic matrix A. Click "multiply" to multiply the colored points by D on the left and A on the right. Note that on both sides, all vectors are "sucked into the 1-eigenspace" (the red line).

Example 6.5.5

Continuing with Red Box example 6.5.1, we can illustrate the Perron–Frobenius theorem explicitly. The matrix

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

has characteristic polynomial

$$f(\lambda) = -\lambda^3 + 0.12\lambda - 0.02 = -(\lambda - 1)(\lambda + 0.2)(\lambda - 0.1).$$

Notice that 1 is strictly greater in absolute value than the other eigenvalues, and that it has algebraic (hence, geometric) multiplicity 1. We compute eigenvectors for the eigenvalues $1, -0.2, 0.1$to be, respectively,

$$u_1 = \begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix} \qquad u_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \qquad u_3 = \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix}.$$

The eigenvector u_1 necessarily has positive entries; the steady-state vector is

$$w = \frac{1}{7+6+5}u_1 = \frac{1}{18}\begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix}.$$

The eigenvectors u_1, u_2, u_3 form a basis $\mathcal{B}$ for $\mathbb{R}^3$; for any vector $x = a_1u_1 + a_2u_2 + a_3u_3$ in $\mathbb{R}^3$, we have

$$\begin{aligned} Ax &= A(a_1u_1 + a_2u_2 + a_3u_3) \\ &= a_1Au_1 + a_2Au_2 + a_3Au_3 \\ &= a_1u_1 - 0.2a_2u_2 + 0.1a_3u_3. \end{aligned}$$

Iterating multiplication by A in this way, we have

$$A^tx = a_1u_1 - (0.2)^ta_2u_2 + (0.1)^ta_3u_3 \;\to\; a_1u_1$$

as $t \to \infty$. This shows that A^tx approaches a_1u_1, which is an *eigenvector with eigenvalue* 1, as guaranteed by the Perron–Frobenius theorem.

What do the above calculations say about the number of copies of Prognosis Negative in the Atlanta Red Box kiosks? Suppose that the kiosks start with 100 copies of the movie, with 30 copies at kiosk 1, 50 copies at kiosk 2, and 20 copies at kiosk 3. Let $v_0 = (30, 50, 20)$ be the vector describing this state. Then there will be $v_1 = Av_0$ movies in the kiosks the next day, $v_2 = Av_1$ the day after that, and so on. We let $v_t = (x_t, y_t, z_t)$.

t	x_t	y_t	z_t
0	30.000000	50.000000	20.000000
1	39.000000	35.000000	26.000000
2	38.700000	33.500000	27.800000
3	38.910000	33.350000	27.740000
4	38.883000	33.335000	27.782000
5	38.889900	33.333500	27.776600
6	38.888670	33.333350	27.777980
7	38.888931	33.333335	27.777734
8	38.888880	33.333333	27.777786
9	38.888891	33.333333	27.777776
10	38.888889	33.333333	27.777778

(Of course it does not make sense to have a fractional number of movies; the decimals are included here to illustrate the convergence.) The steady-state vector says that eventually, the movies will be distributed in the kiosks according to the percentages

$$w = \frac{1}{18}\begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix} = \begin{pmatrix} 38.888888\% \\ 33.333333\% \\ 27.777778\% \end{pmatrix},$$

which agrees with the above table. Moreover, this distribution is *independent* of the beginning distribution of movies in the kiosks.

Now we turn to visualizing the dynamics of (i.e., repeated multiplication by) the matrix A. This matrix is diagonalizable; we have $A = CDC^{-1}$ for

$$C = \begin{pmatrix} 7 & -1 & 1 \\ 6 & 0 & -3 \\ 5 & 1 & 2 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -.2 & 0 \\ 0 & 0 & .1 \end{pmatrix}.$$

The matrix D leaves the x-coordinate unchanged, scales the y-coordinate by $-1/5$, and scales the z-coordinate by $1/10$. Repeated multiplication by D makes the y- and z-coordinates very small, so it "sucks all vectors into the x-axis."

The matrix A does the same thing as D, but with respect to the coordinate system defined by the columns u_1, u_2, u_3 of C. This means that A "sucks all vectors into the 1-eigenspace", without changing the sum of the entries of the vectors.

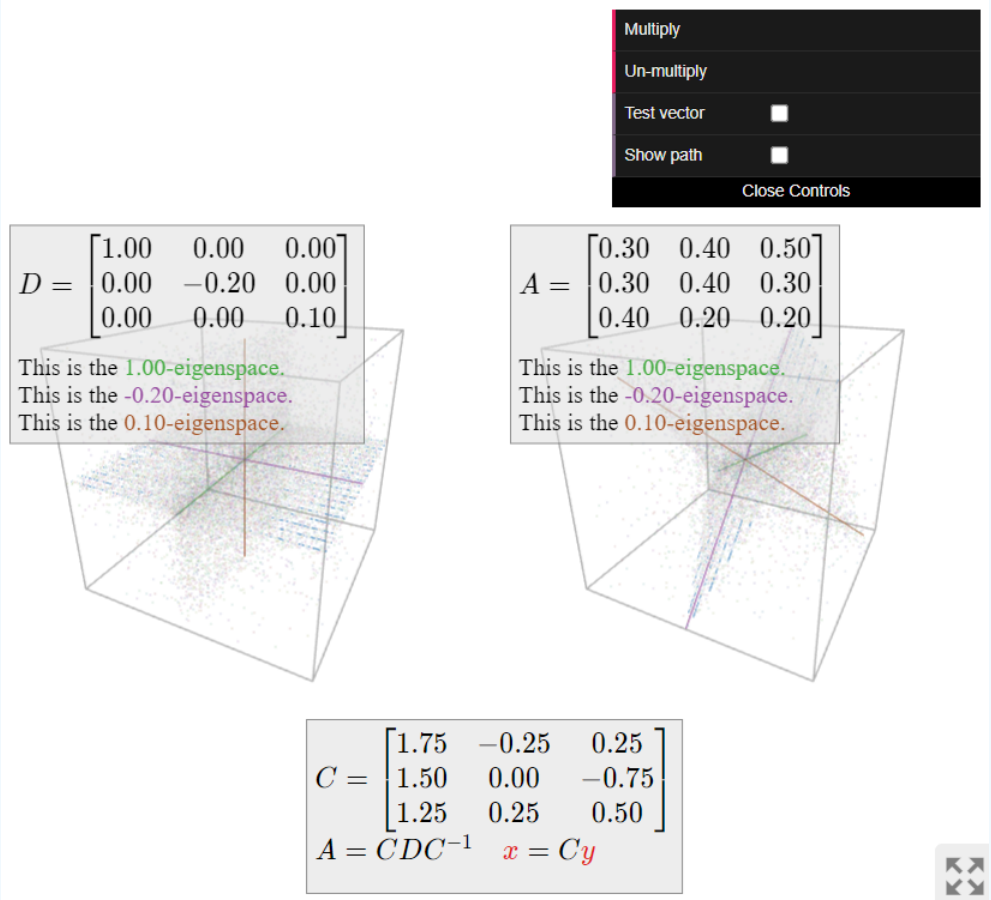


Figure 6.5.3: Dynamics of the stochastic matrix A. Click "multiply" to multiply the colored points by D on the left and A on the right. Note that on both sides, all vectors are "sucked into the 1-eigenspace" (the green line). (We have scaled C by $1/4$ so that vectors have roughly the same size on the right and the left. The "jump" that happens when you press "multiply" is a negation of the $-.2$-eigenspace, which is not animated.)

The picture of a positive stochastic matrix is always the same, whether or not it is diagonalizable: *all vectors are "sucked into the 1-eigenspace,"* which is a line, without changing the sum of the entries of the vectors. This is the geometric content of the Perron–Frobenius theorem.

Google's PageRank Algorithm

Internet searching in the 1990s was very inefficient. Yahoo or AltaVista would scan pages for your search text, and simply list the results with the most occurrences of those words. Not surprisingly, the more unsavory websites soon learned that by putting the words "Alanis Morissette" a million times in their pages, they could show up first every time an angsty teenager tried to find *Jagged Little Pill* on Napster.

Larry Page and Sergey Brin invented a way to rank pages by *importance*. They founded Google based on their algorithm. Here is roughly how it works.

Each web page has an associated importance, or ***rank***. This is a positive number. This rank is determined by the following rule.

Definition 6.5.4: The Importance Rule

If a page P links to n other pages $Q_1, Q_2, \ldots, Q_n$, then each page Q_i inherits $\frac{1}{n}$ of P's importance.

In practice, this means:

- If a very important page links to your page (and not to a zillion other ones as well), then your page is considered important.
- If a zillion unimportant pages link to your page, then your page is still important.
- If only one unknown page links to yours, your page is not important.

Alternatively, there is the *random surfer interpretation.* A "random surfer" just sits at his computer all day, randomly clicking on links. The pages he spends the most time on should be the most important. So, the important (high-ranked) pages are those where a random surfer will end up most often. This measure turns out to be equivalent to the rank.

Definition 6.5.5: The Importance Matrix

Consider an internet with n pages. The ***importance matrix*** is the $n \times n$ matrix A whose i, j-entry is the importance that page j passes to page i.

Observe that the importance matrix is a stochastic matrix, assuming every page contains a link: if page i has m links, then the ith column contains the number $1/m, m$ times, and the number zero in the other entries.

Example 6.5.6

Consider the following internet with only four pages. Links are indicated by arrows.

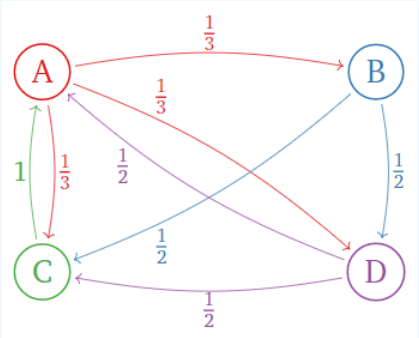


Figure 6.5.4

The importance rule says:

- Page A has 3 links, so it passes $\frac{1}{3}$ of its importance to pages B, C, D.
- Page B has 2 links, so it passes $\frac{1}{2}$ of its importance to pages C, D.
- Page C has one link, so it passes all of its importance to page A.
- Page D has 2 links, so it passes $\frac{1}{2}$ of its importance to pages A, C.

In terms of matrices, if $v = (a, b, c, d)$ is the vector containing the ranks a, b, c, d of the pages A, B, C, D, then

$$\begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} c + \frac{1}{2}d \\ \frac{1}{3}a \\ \frac{1}{3}a + \frac{1}{2}b + \frac{1}{2}d \\ \frac{1}{3}a + \frac{1}{2}b \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}.$$

The matrix on the left is the ***importance matrix***, and the final equality expresses the importance rule.

The above example illustrates the key observation.

Key Observation 6.5.1

The rank vector is an eigenvector of the importance matrix with eigenvalue 1.

In light of the key observation, we would like to use the Perron–Frobenius theorem to find the rank vector. Unfortunately, the importance matrix is not always a *positive* stochastic matrix.

Example 6.5.7: A page with no links

Consider the following internet with three pages:

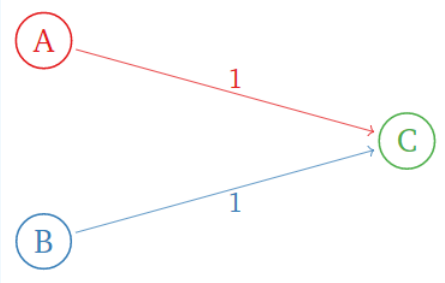


Figure 6.5.5

The importance matrix is

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}.$$

This has characteristic polynomial

$$f(\lambda) = \det \begin{pmatrix} -\lambda & 0 & 0 \\ 0 & -\lambda & 0 \\ 1 & 1 & -\lambda \end{pmatrix} = -\lambda^3.$$

So 1 is not an eigenvalue at all: there is no rank vector! The importance matrix is not stochastic because the page C has no links.

Example 6.5.8: Disconnected Internet

Consider the following internet:

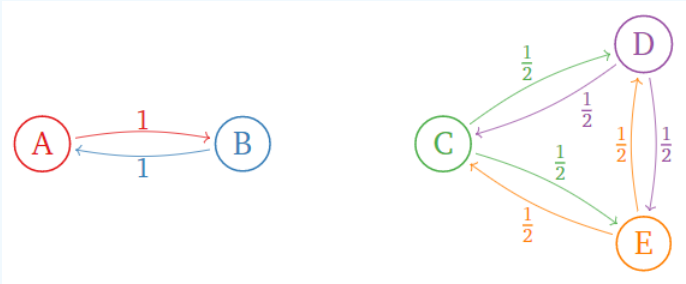


Figure 6.5.6

The importance matrix is

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

This has linearly independent eigenvectors

$$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

both with eigenvalue 1. So there is more than one rank vector in this case. Here the importance matrix is stochastic, but not positive.

Here is Page and Brin's solution. First we fix the importance matrix by replacing each zero column with a column of $1/n$s, where n is the number of pages:

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \quad \text{becomes} \quad A' = \begin{pmatrix} 0 & 0 & 1/3 \\ 0 & 0 & 1/3 \\ 1 & 1 & 1/3 \end{pmatrix}.$$

The ***modified importance matrix*** A' is always stochastic.

Now we choose a number p in $(0, 1)$, called the ***damping factor***. (A typical value is $p = 0.15$.)

Definition 6.5.6: The Google Matrix

Let A be the importance matrix for an internet with n pages, and let A' be the modified importance matrix. The ***Google Matrix*** is the matrix

$$M = (1-p) \cdot A' + p \cdot B \quad \text{where} \quad B = \frac{1}{n} \begin{pmatrix} 1 & 1 & \cdot & 1 \\ 1 & 1 & \cdot & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}.$$

In the random surfer interpretation, this matrix M says: with probability p, our surfer will surf to a completely random page; otherwise, he'll click a random link on the current page, unless the current page has no links, in which case he'll surf to a completely random page in either case. The reader can verify the following important fact.

Fact 6.5.2

The Google Matrix is a positive stochastic matrix.

If we declare that the ranks of all of the pages must sum to 1, then we find:

Note 6.5.3: The 25 Billion Dollar Eigenvector

The PageRank vector is the steady state of the Google Matrix.

This exists and has positive entries by the Perron–Frobenius theorem. The hard part is calculating it: in real life, the Google Matrix has zillions of rows.

Example 6.5.9: PageRank Vector

What is the PageRank vector for the following internet? (Use the damping factor $p = 0.15$.)

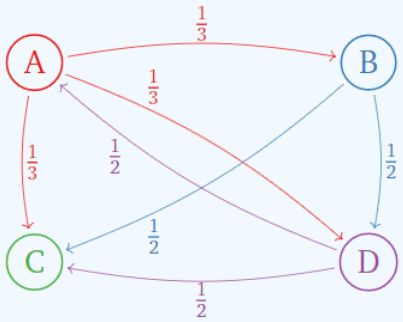


Figure 6.5.7

Which page is the most important? Which is the least important?

Solution

First we compute the modified importance matrix:

$$A = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix} \xrightarrow{\text{modify}} A' = \begin{pmatrix} 0 & 0 & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{4} & 0 \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{4} & 0 \end{pmatrix}$$

Choosing the damping factor $p = 0.15$, the Google Matrix is

$$M = 0.85 \cdot \begin{pmatrix} 0 & 0 & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{4} & 0 \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{4} & 0 \end{pmatrix} + 0.15 \cdot \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$$

$$\approx \begin{pmatrix} 0.0375 & 0.0375 & 0.2500 & 0.4625 \\ 0.3208 & 0.0375 & 0.2500 & 0.0375 \\ 0.3208 & 0.4625 & 0.2500 & 0.4625 \\ 0.3208 & 0.4625 & 0.2500 & 0.0375 \end{pmatrix}.$$

The PageRank vector is the steady state:

$$w \approx \begin{pmatrix} 0.2192 \\ 0.1752 \\ 0.3558 \\ 0.2498 \end{pmatrix}$$

This is the PageRank:

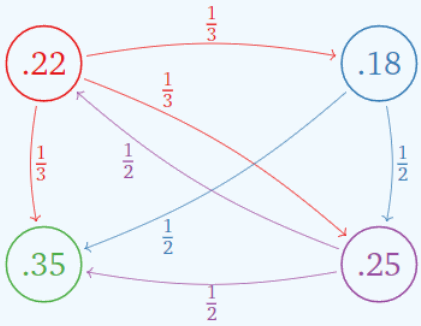


Figure 6.5.8

Page C is the most important, with a rank of 0.558, and page B is the least important, with a rank of 0.1752.

This page titled 6.5: Stochastic Matrices is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **5.6: Stochastic Matrices** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

6.6: Similarity

Objectives

1. Learn to interpret similar matrices geoemetrically.
2. Understand the relationship between the eigenvalues, eigenvectors, and characteristic polynomials of similar matrices.
3. *Recipe:* compute Ax in terms of B, C for $A = CBC^{-1}$.
4. *Picture:* the geometry of similar matrices.
5. *Vocabulary word:* ***similarity***.

Some matrices are easy to understand. For instance, a diagonal matrix

$$D = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

just scales the coordinates of a vector: $D\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x \\ y/2 \end{pmatrix}$. The purpose of most of the rest of this chapter is to understand complicated-looking matrices by analyzing to what extent they "behave like" simple matrices. For instance, the matrix

$$A = \frac{1}{10}\begin{pmatrix} 11 & 6 \\ 9 & 14 \end{pmatrix}$$

has eigenvalues 2 and $1/2$, with corresponding eigenvectors $v_1 = \begin{pmatrix} 2/3 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Notice that

$$\begin{aligned} D(xe_1 + ye_2) &= xDe_1 + yDe_2 = 2xe_1 - \frac{1}{2}ye_2 \\ A(xv_1 + yv_2) \ &= xAv_1 + yAv_2 = 2xv_1 - \frac{1}{2}yv_2. \end{aligned}$$

Using $v_1,\ v_2$ instead of the usual coordinates makes A "behave" like a diagonal matrix.

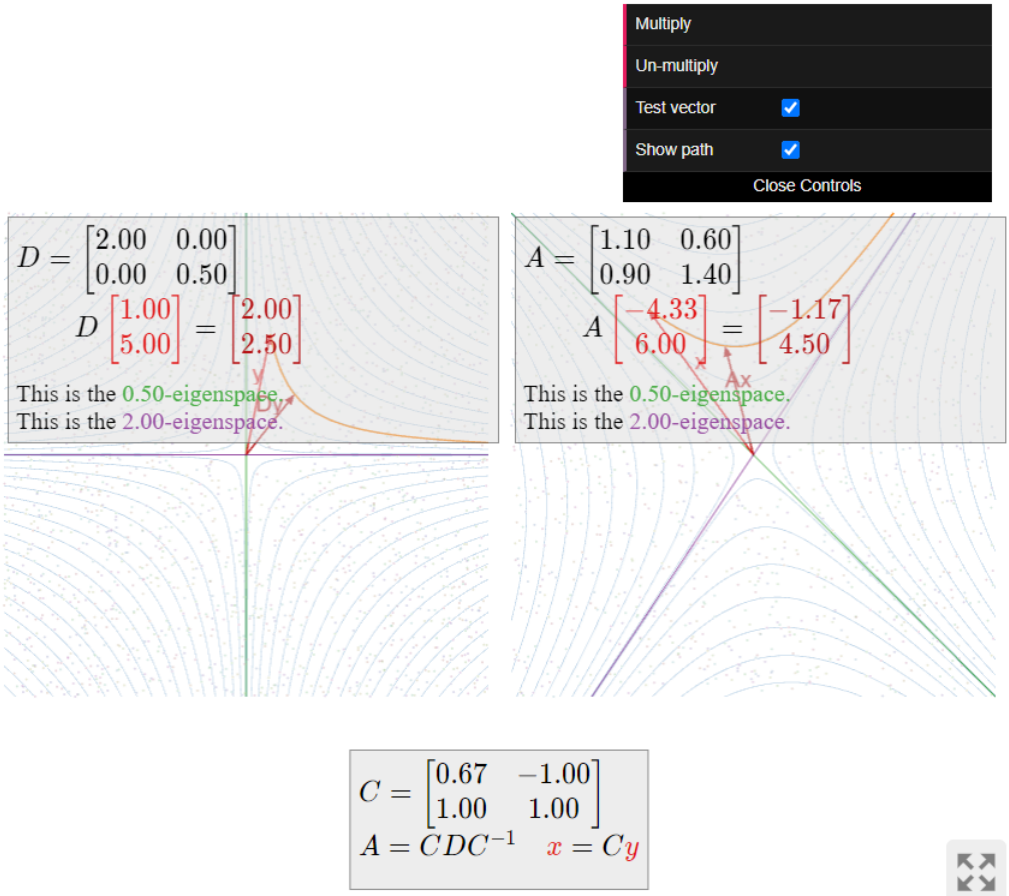


Figure 6.6.1: The matrices A and D behave similarly. Click "multiply" to multiply the colored points by D on the left and A on the right. (We will see in Section 5.4 why the points follow hyperbolic paths.)

The other case of particular importance will be matrices that "behave" like a rotation matrix: indeed, this will be crucial for understanding Section 5.5 geometrically. See Note 6.6.3.

In this section, we study in detail the situation when two matrices behave similarly with respect to different coordinate systems. In Section 5.4 and Section 5.5, we will show how to use eigenvalues and eigenvectors to find a simpler matrix that behaves like a given matrix.

Similar Matrices

We begin with the algebraic definition of similarity.

Definition 6.6.1: Similar

Two $n \times n$ matrices A and B are ***similar*** if there exists an invertible $n \times n$ matrix C such that $A = CBC^{-1}$.

Example 6.6.1

The matrices

$$\begin{pmatrix} -12 & 15 \\ -10 & 13 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}$$

are similar because

$$\begin{pmatrix} -12 & 15 \\ -10 & 13 \end{pmatrix} = \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix}^{-1},$$

as the reader can verify.

Example 6.6.2

The matrices

$$\begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

are not similar. Indeed, the second matrix is the identity matrix I_2, so if C is any invertible 2×2 matrix, then

$$CI_2C^{-1} = CC^{-1} = I_2 \neq \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}.$$

As in the above example, one can show that I_n is the only matrix that is similar to I_n, and likewise for any scalar multiple of I_n.

Note 6.6.1

Similarity is unrelated to row equivalence. Any invertible matrix is row equivalent to I_n, but I_n is the only matrix similar to I_n. For instance,

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

are row equivalent but not similar.

As suggested by its name, similarity is what is called an *equivalence relation*. This means that it satisfies the following properties.

Proposition 6.6.1

Let A, B, and C be $n \times n$ matrices.

1. **Reflexivity:** A is similar to itself.
2. **Symmetry:** if A is similar to B, then B is similar to A.
3. **Transitivity:** if A is similar to B and B is similar to C, then A is similar to C.

Proof

1. Taking $C = I_n = I_n^{-1}$, we have $A = I_n A I_n^{-1}$.
2. Suppose that $A = CBC^{-1}$. Multiplying both sides on the left by C^{-1} and on the right by C gives
$$C^{-1}AC = C^{-1}(CBC^{-1})C = B.$$
Since $(C^{-1})^{-1} = C$, we have $B = C^{-1}A(C^{-1})^{-1}$, so that B is similar to A.
3. Suppose that $A = DBD^{-1}$ and $B = ECE^{-1}$. Subsituting for B and remembering that $(DE)^{-1} = E^{-1}D^{-1}$, we have
$$A = D(ECE^{-1})D^{-1} = (DE)C(DE)^{-1},$$
which shows that A is similar to C.

Example 6.6.3

The matrices

$$\begin{pmatrix} -12 & 15 \\ -10 & 13 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}$$

are similar, as we saw in Example 6.6.1. Likewise, the matrices

$$\begin{pmatrix} -12 & 15 \\ -10 & 13 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -12 & 5 \\ -30 & 13 \end{pmatrix}$$

are similar because

$$\begin{pmatrix} -12 & 5 \\ -30 & 13 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} -12 & 15 \\ -10 & 13 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 2 & 1 \end{pmatrix}^{-1}.$$

It follows that

$$\begin{pmatrix} -12 & 5 \\ -30 & 13 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}$$

are similar to each other.

We conclude with an observation about similarity and powers of matrices.

Fact 6.6.1

Let $A = CBC^{-1}$. Then for any $n \geq 1$, we have

$$A^n = CB^nC^{-1}.$$

Proof

First note that

$$A^2 = AA = (CBC^{-1})(CBC^{-1}) = CB(C^{-1}C)BC^{-1} = CBI_nBC^{-1} = CB^2C^{-1}.$$

Next we have

$$A^3 = A^2A = (CB^2C^{-1})(CBC^{-1}) = CB^2(C^{-1}C)BC^{-1} = CB^3C^{-1}.$$

The pattern is clear.

Example 6.6.4

Compute A^{100}, where

$$A = \begin{pmatrix} 5 & 13 \\ -2 & -5 \end{pmatrix} = \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix}^{-1}.$$

Solution

By the fact 6.6.1, we have

$$A^{100} = \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^{100} \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix}^{-1}.$$

The matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is a counterclockwise rotation by 90°. If we rotate by 90° four times, then we end up where we started. Hence rotating by 90° one hundred times is the identity transformation, so

$$A^{100} = \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Geometry of Similar Matrices

Similarity is a very interesting construction when viewed geometrically. We will see that, roughly, *similar matrices do the same thing in different coordinate systems.* The reader might want to review $\mathcal{B}$-coordinates and nonstandard coordinate grids in Section 2.8 before reading this subsection.

By conditions 4 and 5 of the invertible matrix theorem in Section 5.1, an $n \times n$ matrix C is invertible if and only if its columns $v_1, v_2, \cdots, v_n$ form a basis for $\mathbb{R}^n$. This means we can speak of the $\mathcal{B}$-coordinates of a vector in $\mathbb{R}^n$, where $\mathcal{B}$ is the basis of

columns of C. Recall that

$$[x]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \quad \text{means} \quad x = c_1v_1 + c_2v_2 + \cdots + c_nv_n = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}.$$

Since C is the matrix with columns $v_1, v_2, \cdots, v_n$, this says that $x = C[x]_{\mathcal{B}}$. Multiplying both sides by C^{-1} gives $[x]_{\mathcal{B}} = C^{-1}x$. To summarize:

Note 6.6.2

Let C be an invertible $n \times n$ matrix with columns $v_1, v_2, \cdots, v_n$, and let $\mathcal{B} = \{v_1, v_2, \cdots, v_n\}$, a basis for $\mathbb{R}^n$. Then for any x in $\mathbb{R}^n$, we have

$$C[x]_{\mathcal{B}} = x \quad \text{and} \quad C^{-1}x = [x]_{\mathcal{B}}.$$

This says that C changes from the $\mathcal{B}$-coordinates to the usual coordinates, and C^{-1} changes from the usual coordinates to the $\mathcal{B}$-coordinates.

Suppose that $A = CBC^{-1}$. The above observation gives us another way of computing Ax for a vector x in $\mathbb{R}^n$. Recall that $CBC^{-1}x = C(B(C^{-1}x))$, so that multiplying CBC^{-1} by x means *first* multiplying by C^{-1}, *then* by B, *then* by C. See Example 3.4.10 in Section 3.4.

Recipe: Computing Ax in terms of B.

Suppose that $A = CBC^{-1}$, where C is an invertible matrix with columns $v_1, v_2, \cdots, v_n$. Let $\mathcal{B} = \{v_1, v_2, \cdots, v_n\}$, a basis for $\mathbb{R}^n$. Let x be a vector in $\mathbb{R}^n$. To compute Ax, one does the following:

1. Multiply x by C^{-1}, which changes to the $\mathcal{B}$-coordinates: $[x]_{\mathcal{B}} = C^{-1}x$.
2. Multiply this by B: $B[x]_{\mathcal{B}} = BC^{-1}x$.
3. Interpreting this vector as a $\mathcal{B}$-coordinate vector, we multiply it by C to change back to the usual coordinates: $Ax = CBC^{-1}x = CB[x]_{\mathcal{B}}$.

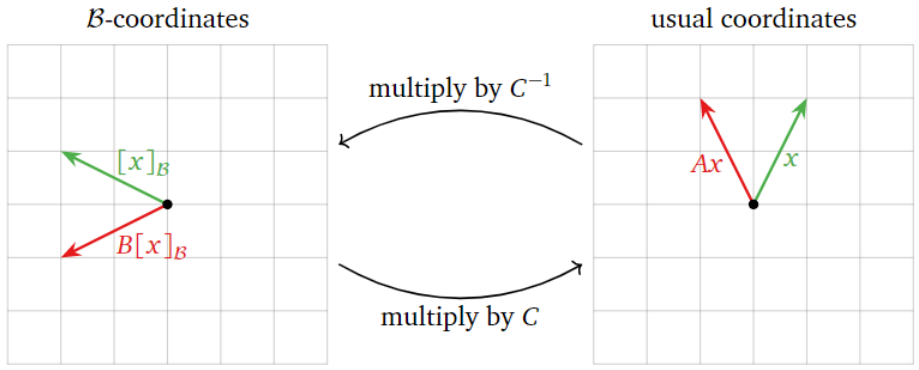


Figure 6.6.2

To summarize: if $A = CBC^{-1}$, then A and B do the same thing, only in different coordinate systems.

The following example is the heart of this section.

Example 6.6.5

Consider the matrices

$$A = \begin{pmatrix} 1/2 & 3/2 \\ 3/2 & 1/2 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \quad C = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

One can verify that $A = CBC^{-1}$: see Example 5.4.1 in Section 5.4. Let $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, the columns of C, and let $\mathcal{B} = \{v_1,\ v_2\}$, a basis of $\mathbb{R}^2$.

The matrix B is diagonal: it scales the x-direction by a factor of 2 and the y-direction by a factor of -1.

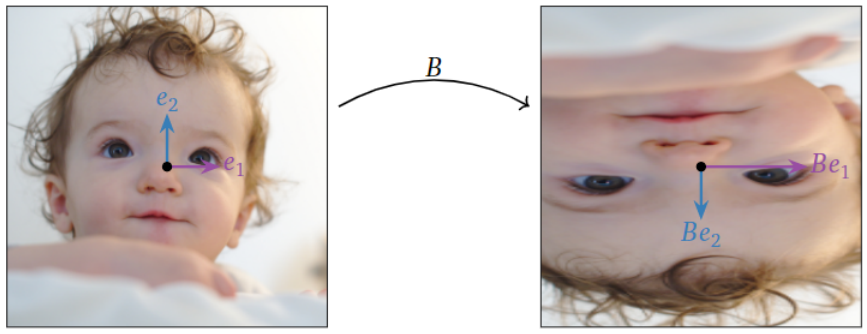


Figure 6.6.3

To compute Ax, first we multiply by C^{-1} to find the $\mathcal{B}$-coordinates of x, then we multiply by B, then we multiply by C again. For instance, let $x = \begin{pmatrix} 0 \\ -2 \end{pmatrix}$.

1. We see from the $\mathcal{B}$-coordinate grid below that $x = -v_1 + v_2$. Therefore, $C^{-1}x = [x]_{\mathcal{B}} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$..
2. Multiplying by B scales the coordinates: $B[x]_{\mathcal{B}} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$.
3. Interpreting $\begin{pmatrix} -2 \\ -1 \end{pmatrix}$ as a $\mathcal{B}$-coordinate vector, we multiply by C to get

$$Ax = C\begin{pmatrix} -2 \\ -1 \end{pmatrix} = -2v_1 - v_2 = \begin{pmatrix} -3 \\ -1 \end{pmatrix}.$$

Of course, this vector lies at $(-2, -1)$ on the $\mathcal{B}$-coordinate grid.

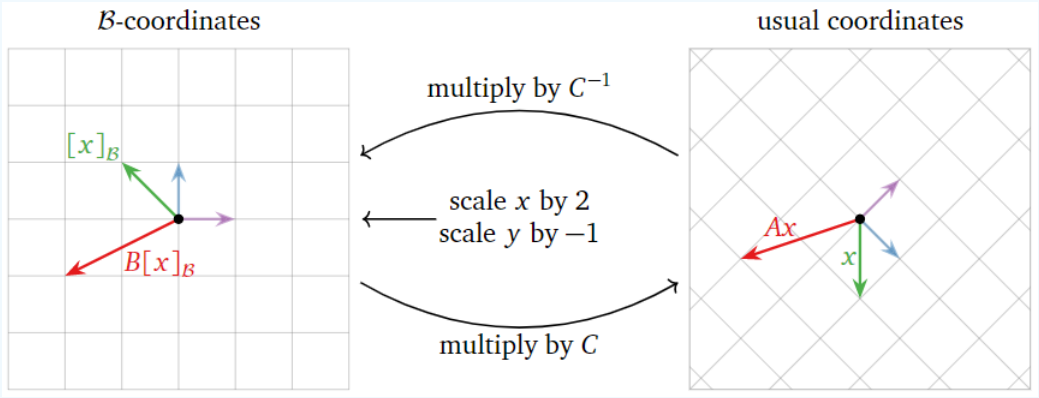


Figure 6.6.4

Now let $x = \frac{1}{2}\begin{pmatrix} 5 \\ -3 \end{pmatrix}$.

1. We see from the $\mathcal{B}$-coordinate grid that $x = \frac{1}{2}v_1 + 2v_2$. Therefore, $C^{-1}x = [x]_{\mathcal{B}} = \begin{pmatrix} 1/2 \\ 2 \end{pmatrix}$.
2. Multiplying by B scales the coordinates: $B[x]_{\mathcal{B}} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$.
3. Interpreting $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ as a $\mathcal{B}$-coordinate vector, we multiply by C to get

$$Ax = C\begin{pmatrix}1\\-2\end{pmatrix} = v_1 - 2v_2 = \begin{pmatrix}-1\\3\end{pmatrix}.$$

This vector lies at $(1, -2)$ on the $\mathcal{B}$-coordinate grid.

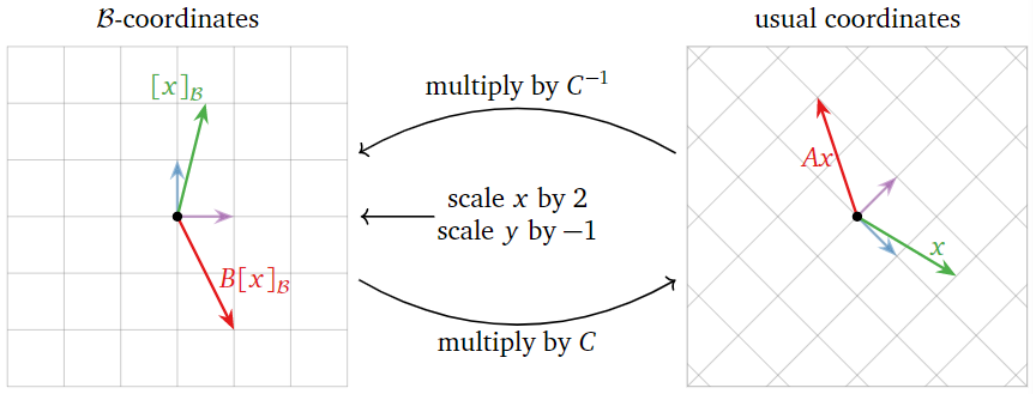


Figure 6.6.5

To summarize:

- B scales the e_1-direction by 2 and the e_2-direction by -1.
- A scales the v_1-direction by 2 and the v_2-direction by -1.

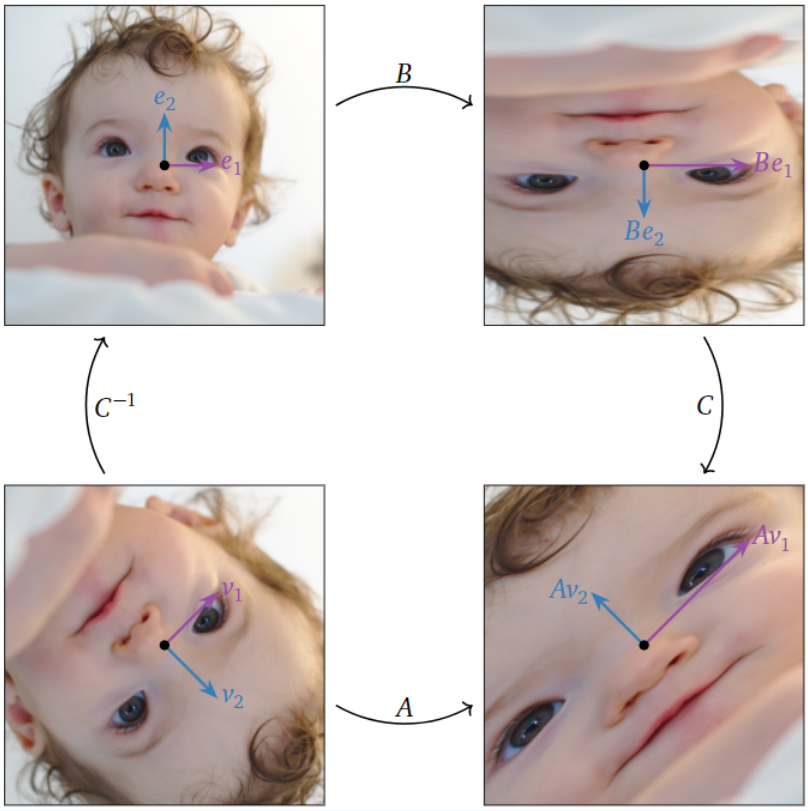


Figure 6.6.6

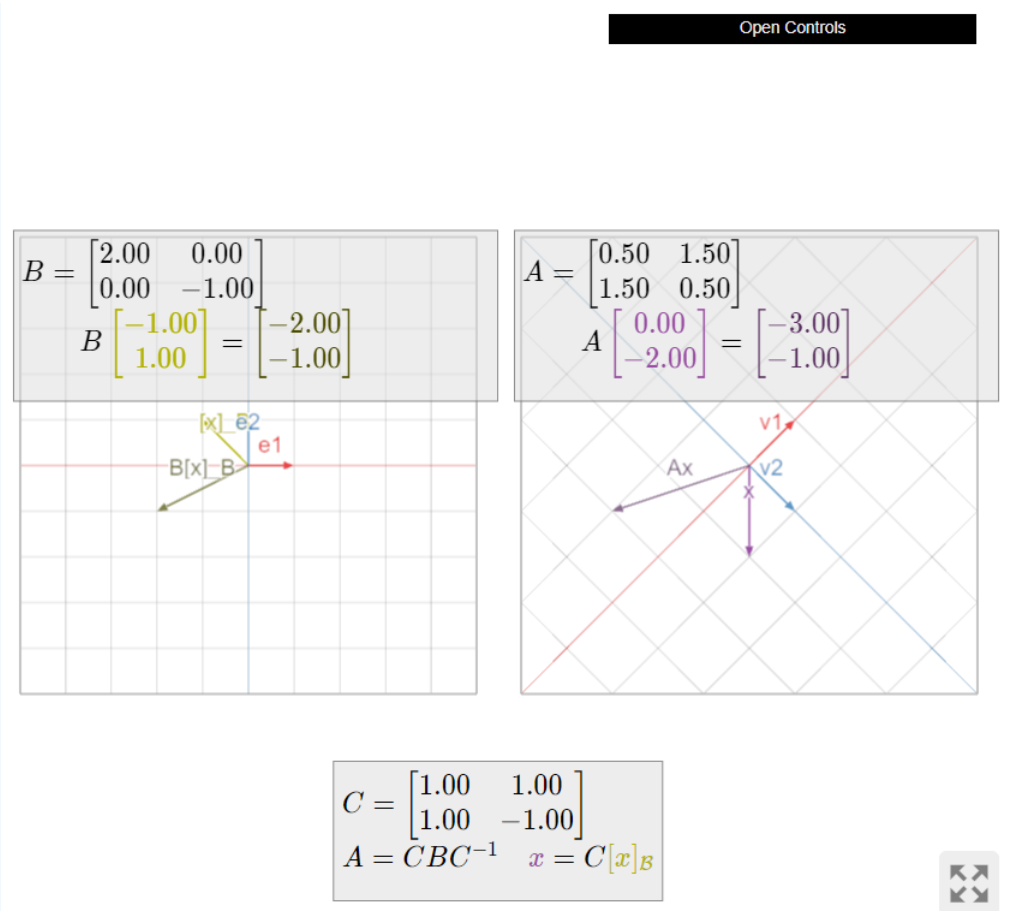


Figure 6.6.7: The geometric relationship between the similar matrices A and B acting on $\mathbb{R}^2$. Click and drag the heads of x and $[x]_{\mathcal{B}}$. Study this picture until you can reliably predict where the other three vectors will be after moving one of them: this is the essence of the geometry of similar matrices.

Example 6.6.6: Interactive: Another matrix similar to B

Consider the matrices

$$A' = \frac{1}{5}\begin{pmatrix} -8 & -9 \\ 6 & 13 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \quad C' = \frac{1}{2}\begin{pmatrix} -1 & -3 \\ 2 & 1 \end{pmatrix}.$$

Then $A' = C'B(C')^{-1}$, as one can verify. Let $v_1' = \frac{1}{2}\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ and $v_2' = \frac{1}{2}\begin{pmatrix} -3 \\ 1 \end{pmatrix}$, the columns of C', and let $\mathcal{B}' = \{v_1', v_2'\}$. Then A' does the same thing as B, as in the previous example 6.6.5, except A' uses the $\mathcal{B}'$-coordinate system. In other words:

- B scales the e_1-direction by 2 and the e_2-direction by -1.
- A' scales the v_1'-direction by 2 and the v_2'-direction by -1.

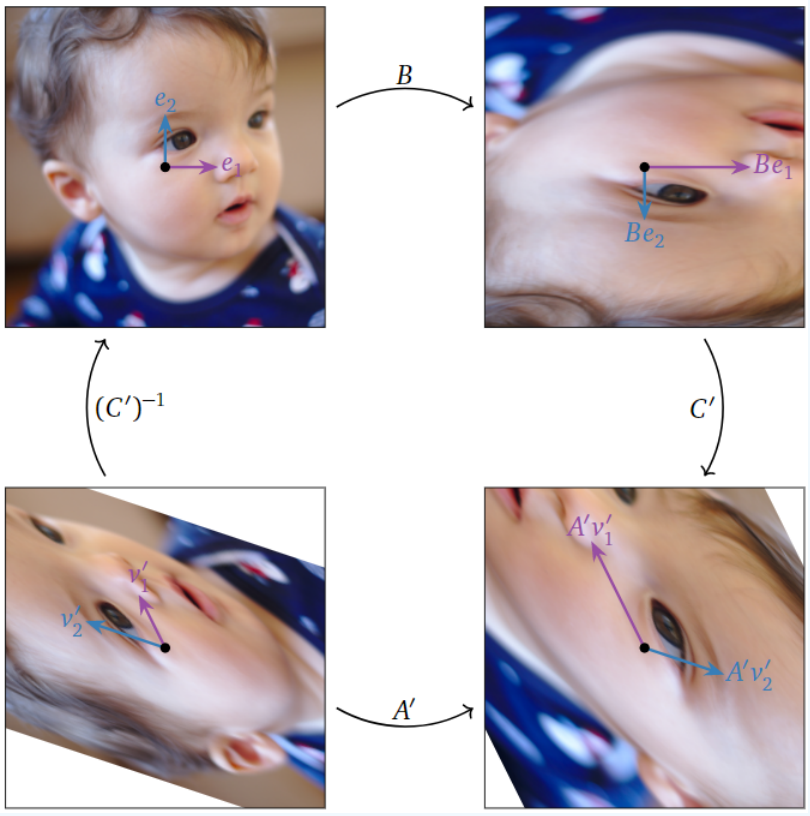

e_2
e_1
B
Be_1
Be_2
$(C')^{-1}$
C'
v_1'
v_2'
A'
$A'v_1'$
$A'v_2'$

Figure 6.6.8

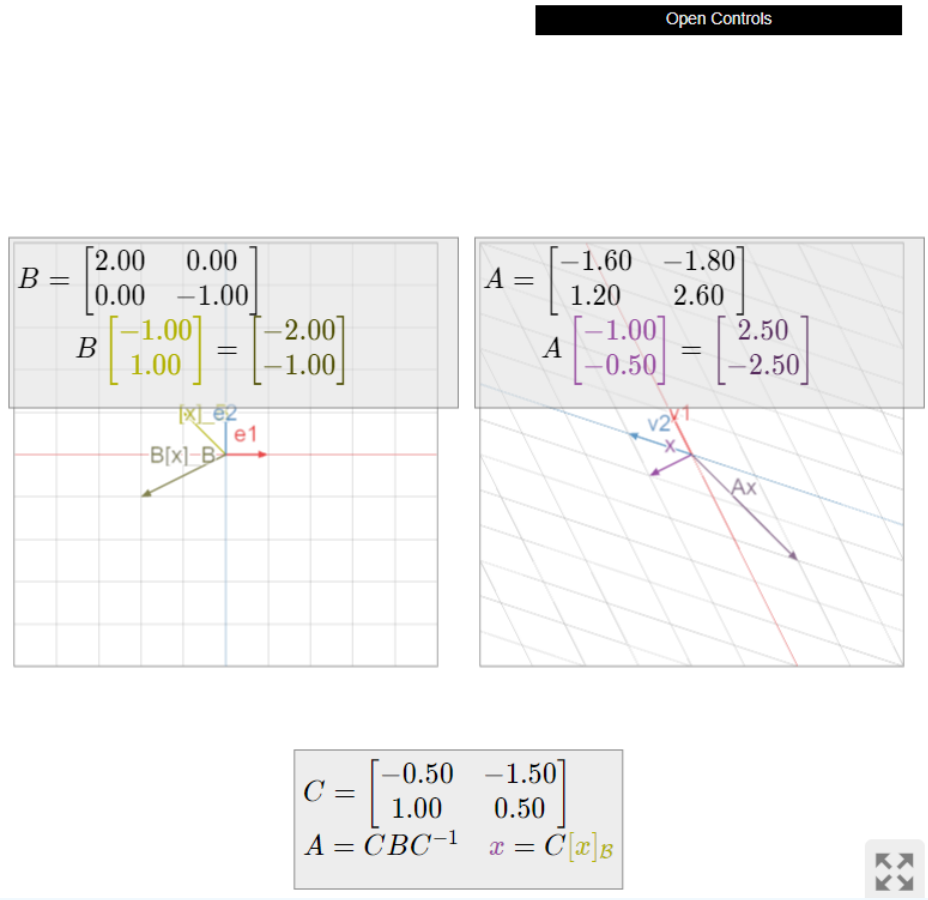

Open Controls
$B = \begin{bmatrix} 2.00 & 0.00 \\ 0.00 & -1.00 \end{bmatrix}$
$B \begin{bmatrix} -1.00 \\ 1.00 \end{bmatrix} = \begin{bmatrix} -2.00 \\ -1.00 \end{bmatrix}$
$A = \begin{bmatrix} -1.60 & -1.80 \\ 1.20 & 2.60 \end{bmatrix}$
$A \begin{bmatrix} -1.00 \\ -0.50 \end{bmatrix} = \begin{bmatrix} 2.50 \\ -2.50 \end{bmatrix}$
[x]_e2
e1
B[x]_B
v1
v2
x
Ax
$C = \begin{bmatrix} -0.50 & -1.50 \\ 1.00 & 0.50 \end{bmatrix}$
$A = CBC^{-1}$
$x = C[x]_B$

Figure 6.6.9: The geometric relationship between the similar matrices A' and B acting on $\mathbb{R}^2$. Click and drag the heads of x and $[x]_{\mathcal{B}'}$.

Example 6.6.7: A matrix similar to a rotation matrix

Consider the matrices

$$A = \frac{1}{6}\begin{pmatrix} 7 & -17 \\ 5 & -7 \end{pmatrix} \quad B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad C = \begin{pmatrix} 2 & -1/2 \\ 1 & 1/2 \end{pmatrix}.$$

One can verify that $A = CBC^{-1}$. Let $v_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $v_2 = \frac{1}{2}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$, the columns of C, and let $\mathcal{B} = \{v_1, v_2\}$, a basis of $\mathbb{R}^2$.

The matrix B rotates the plane counterclockwise by 90°.

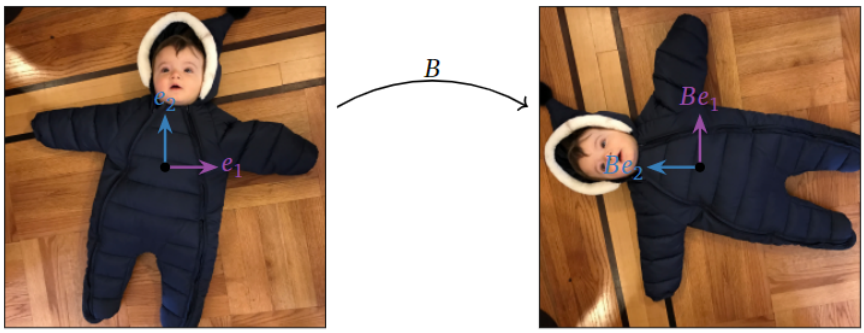


Figure 6.6.10

To compute Ax, first we multiply by C^{-1} to find the $\mathcal{B}$-coordinates of x, then we multiply by B, then we multiply by C again. For instance, let $x = \frac{3}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

1. We see from the $\mathcal{B}$-coordinate grid below that $x = v_1 + v_2$. Therefore, $C^{-1}x = [x]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.
2. Multiplying by B rotates by 90°: $B[x]_{\mathcal{B}} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.
3. Interpreting $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ as a $\mathcal{B}$-coordinate vector, we multiply by C to get

$$Ax = C\begin{pmatrix} -1 \\ 1 \end{pmatrix} = -v_1 + v_2 = \frac{1}{2}\begin{pmatrix} -5 \\ 1 \end{pmatrix}.$$

Of course, this vector lies at $(-1, 1)$ on the $\mathcal{B}$-coordinate grid.

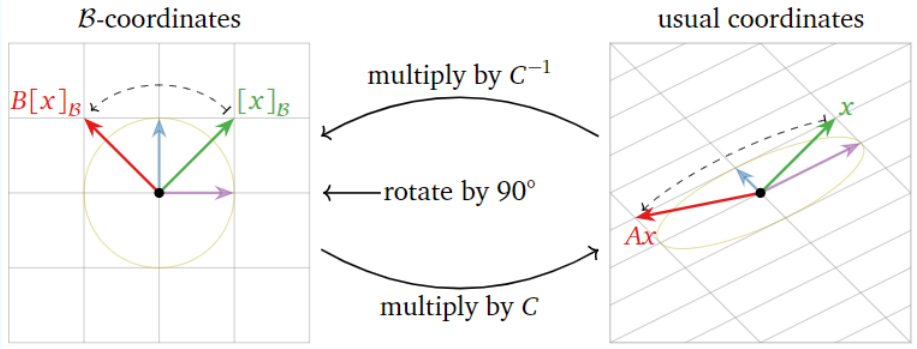


Figure 6.6.11

Now let $x = \frac{1}{2}\begin{pmatrix} -1 \\ -2 \end{pmatrix}$.

1. We see from the $\mathcal{B}$-coordinate grid that $x = -\frac{1}{2}v_1 - v_2$. Therefore, $C^{-1}x = [x]_{\mathcal{B}} = \begin{pmatrix} -1/2 \\ -1 \end{pmatrix}$.
2. Multiplying by B rotates by 90°: $B[x]_{\mathcal{B}} = \begin{pmatrix} 1 \\ -1/2 \end{pmatrix}$.
3. Interpreting $\begin{pmatrix} 1 \\ -1/2 \end{pmatrix}$ as a $\mathcal{B}$-coordinate vector, we multiply by C to get

$$Ax = C\begin{pmatrix} 1 \\ -1/2 \end{pmatrix} = v_1 - \frac{1}{2}v_2 = \frac{1}{4}\begin{pmatrix} 9 \\ 3 \end{pmatrix}.$$

This vector lies at $(1, -\frac{1}{2})$ on the $\mathcal{B}$-coordinate grid.

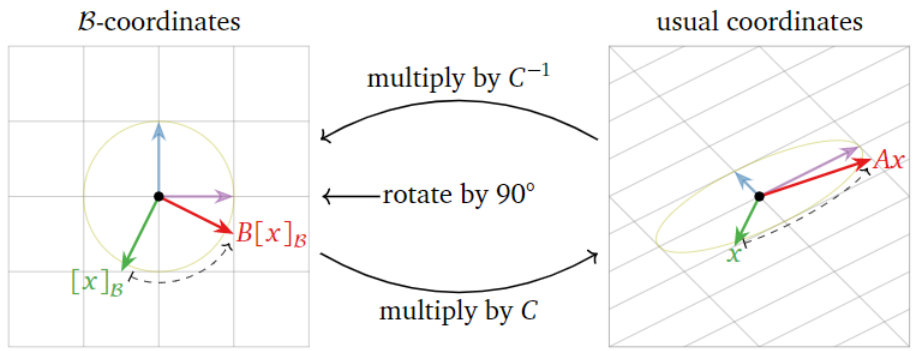


Figure 6.6.12

To summarize:

- B rotates counterclockwise around the circle centered at the origin and passing through e_1 and e_2.
- A rotates counterclockwise around the ellipse centered at the origin and passing through v_1 and v_2.

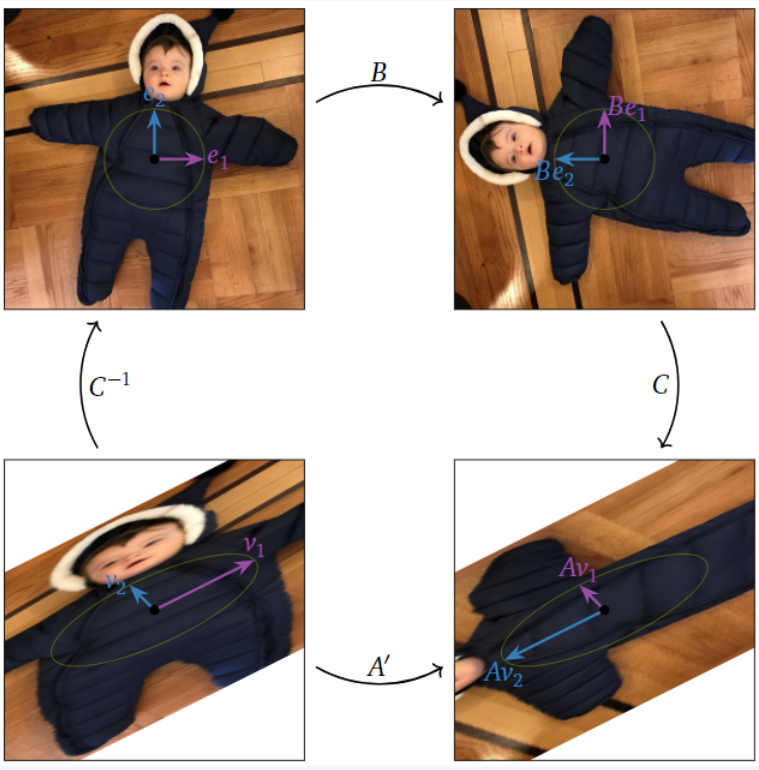


Figure 6.6.13

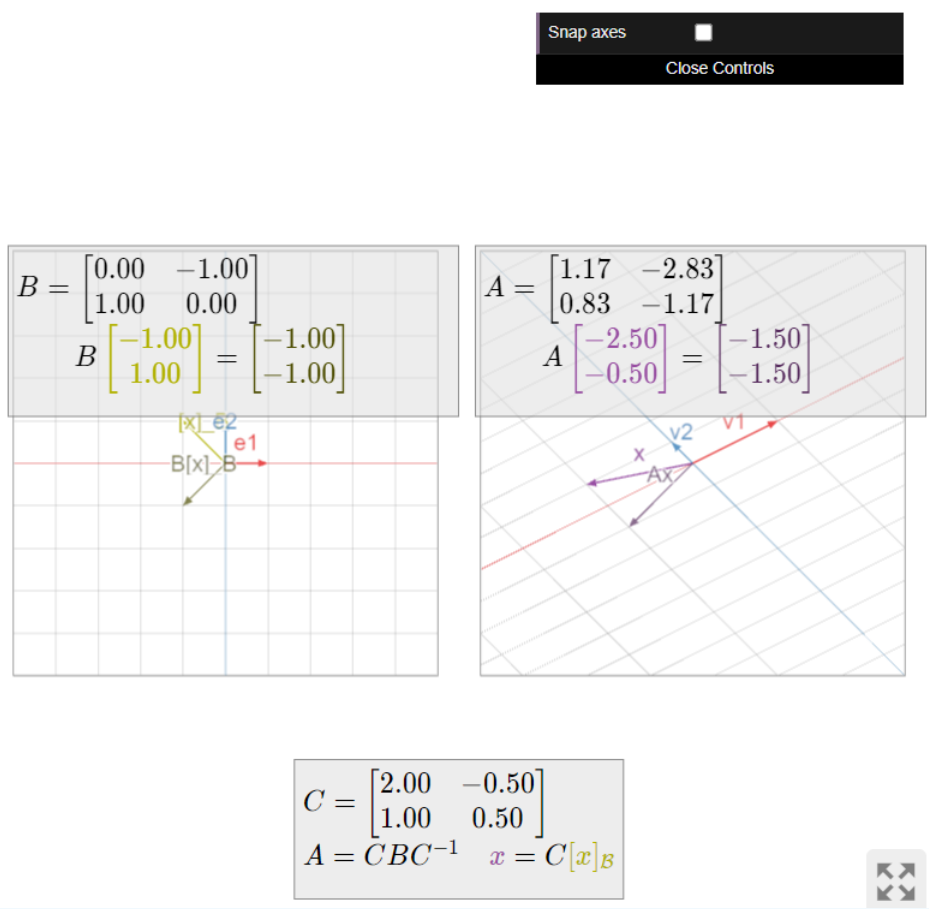


Figure 6.6.14: The geometric relationship between the similar matrices A and B acting on $\mathbb{R}^2$. Click and drag the heads of x and $[x]_{\mathcal{B}}$.

To summarize and generalize the previous example:

Note 6.6.3

Let

$$B = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad C = \begin{pmatrix} | & | \\ v_1 & v_2 \\ | & | \end{pmatrix} \quad A = CBC^{-1},$$

where C is assumed invertible. Then:

- B rotates the plane by an angle of θ around the circle centered at the origin and passing through e_1 and e_2, in the direction from e_1 to e_2.
- A rotates the plane by an angle of θ around the ellipse centered at the origin and passing through v_1 and v_2, in the direction from v_1 to v_2.

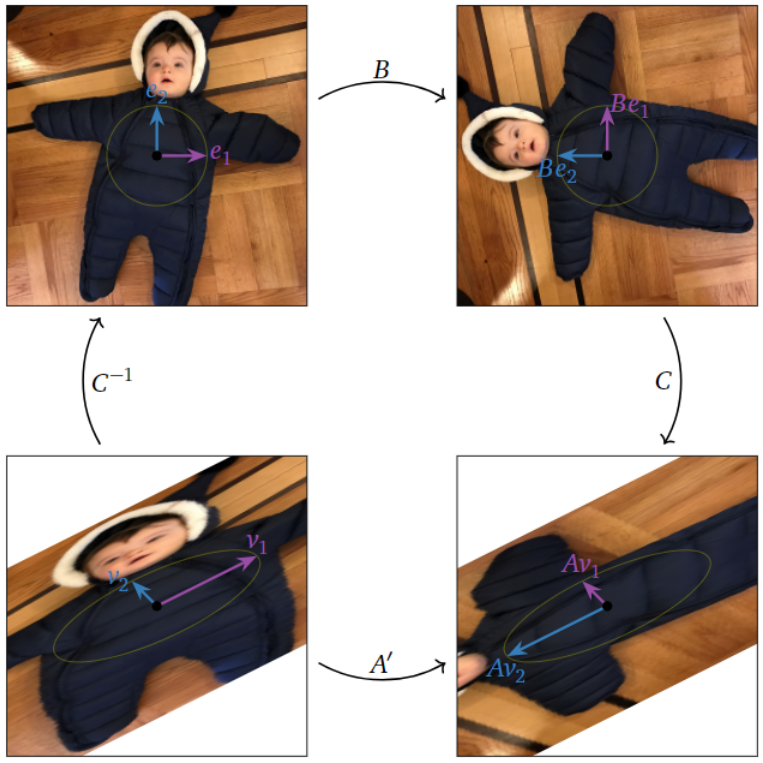


Figure 6.6.15

Example 6.6.8: Interactive: Similar 3×3 matrices

Consider the matrices

$$A = \begin{pmatrix} -1 & 0 & 0 \\ -1 & 0 & 2 \\ -1 & 1 & 1 \end{pmatrix} \quad B = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \quad C = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix}.$$

Then $A = CBC^{-1}$, as one can verify. Let $v_1,\ v_2,\ v_3$ be the columns of C, and let $\mathcal{B} = \{v_1,\ v_2,\ v_3\}$, a basis of $\mathbb{R}^3$. Then A does the same thing as B, except A uses the $\mathcal{B}$-coordinate system. In other words:

- B scales the e_1, e_2-plane by -1 and the e_3-direction by 2.
- A scales the v_1, v_2-plane by -1 and the v_3-direction by 2.

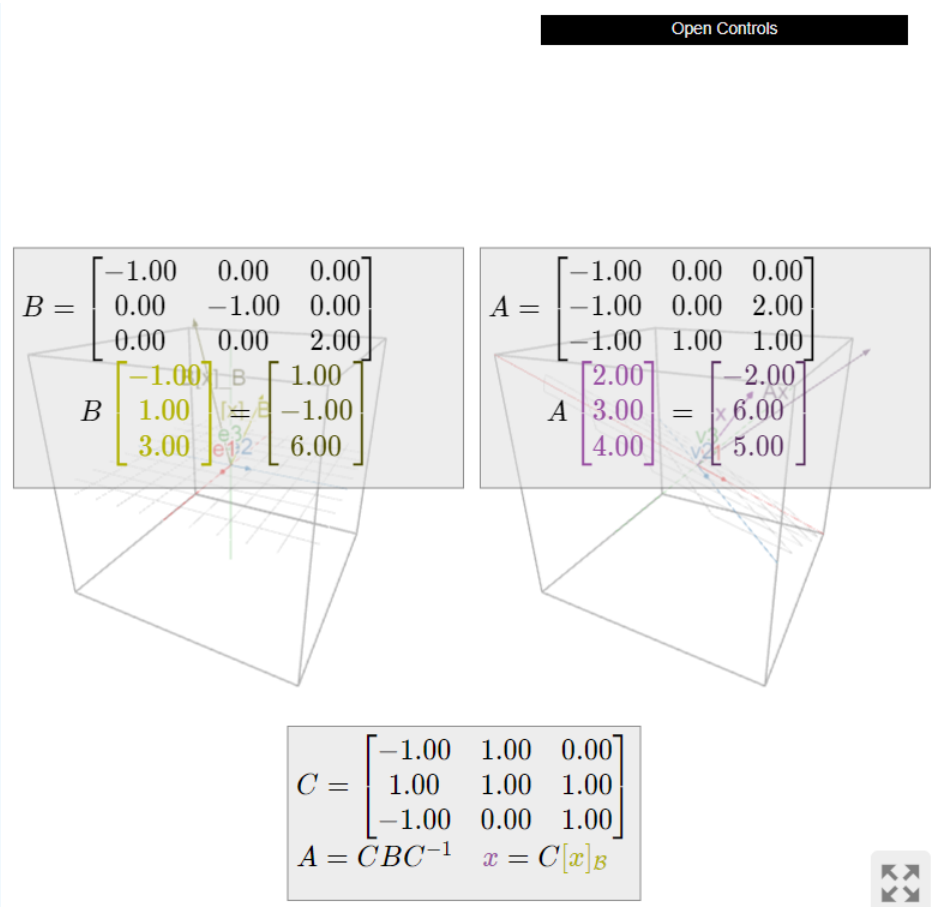


Figure 6.6.16: The geometric relationship between the similar matrices A and B acting on $\mathbb{R}^3$. Click and drag the heads of x and $[x]_{\mathcal{B}}$.

Eigenvalues of Similar Matrices

Since similar matrices behave in the same way with respect to different coordinate systems, we should expect their eigenvalues and eigenvectors to be closely related.

Fact 6.6.2

Similar matrices have the same characteristic polynomial.

Proof

Suppose that $A = CBC^{-1}$, where A, B, C are $n \times n$ matrices. We calculate

$$\begin{aligned} A - \lambda I_n &= CBC^{-1} - \lambda CC^{-1} = CBC^{-1} - C\lambda C^{-1} \\ &= CBC^{-1} - C\lambda I_n C^{-1} = C(B_\lambda I_n)C^{-1}. \end{aligned}$$

Therefore,

$$\det(A - \lambda I_n) = \det(C(B - \lambda I_n)C^{-1}) = \det(C)\det(B - \lambda I_n)\det(C)^{-1} = \det(B - \lambda I_n).$$

Here we have used the multiplicativity property Proposition 4.1.3 in Section 4.1 and its Corollary 4.1.2 in Section 4.1.

Since the eigenvalues of a matrix are the roots of its characteristic polynomial, we have shown:

Note 6.6.4

Similar matrices have the same eigenvalues.

By Theorem 5.2.2 in Section 5.2, similar matrices also have the same trace and determinant.

Note 6.6.5

The converse of fact 6.6.2 is false. Indeed, the matrices

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

both have the characteristic polynomial $f(\lambda) = (\lambda - 1)^2$, but they are not similar, because the only matrix that is similar to I_2 is I_2 itself.

Given that similar matrices have the same eigenvalues, one might guess that they have the same eigen*vectors* as well. Upon reflection, this is not what one should expect: indeed, the eigenvectors should only match up after changing from one coordinate system to another. This is the content of the next fact, remembering that C and C^{-1} change between the usual coordinates and the $\mathcal{B}$-coordinates.

Fact 6.6.3

Suppose that $A = CBC^{-1}$. Then

$$\begin{aligned} v \text{ is an eigenvector of } A &\implies C^{-1}v \text{ is an eigenvector of } B \\ v \text{is an eigenvector of } B &\implies Cv \text{ is an eigenvector of } A. \end{aligned}$$

The eigenvalues of $v/C^{-1}v$ or v/Cv are the same.

Proof

Suppose that v is an eigenvector of A with eigenvalue λ, so that $Av = \lambda v$. Then

$$B(C^{-1}v) = C^{-1}(CBC^{-1}v) = C^{-1}(Av) = C^{-1}\lambda v = \lambda(C^{-1}v),$$

so that $C^{-1}v$ is an eigenvector of B with eigenvalue λ. Likewise if v is an eigenvector of B with eigenvalue λ, then $Bv = \lambda v$, and we have

$$A(Cv) = (CBC^{-1})Cv = CBv = C(\lambda v) = \lambda(Cv),$$

so that Cv is an eigenvalue of A with eigenvalue λ.

Note 6.6.6

If $A = CBC^{-1}$, then C^{-1} takes the λ-eigenspace of A to the λ-eigenspace of B, and C takes the λ-eigenspace of B to the λ-eigenspace of A.

Example 6.6.9

We continue with the above example 6.6.5: let

$$A = \begin{pmatrix} 1/2 & 3/2 \\ 3/2 & 1/2 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \quad C = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

so $A = CBC^{-1}$. Let $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, the columns of C. Recall that:

- B scales the e_1-direction by 2 and the e_2-direction by -1.
- A scales the v_1-direction by 2 and the v_2-direction by -1.

This means that the x-axis is the 2-eigenspace of B, and the y-axis is the -1-eigenspace of B; likewise, the "v_1-axis" is the 2-eigenspace of A, and the "v_2-axis" is the -1-eigenspace of A. This is consistent with the fact 6.6.3, as multiplication by C changes e_1 into $Ce_1 = v_1$ and e_2 into $Ce_2 = v_2$.

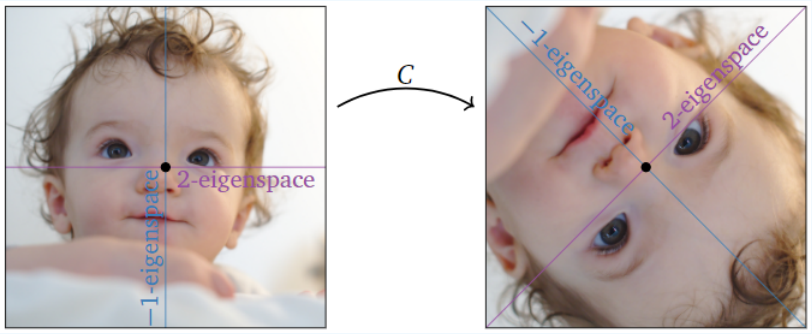


Figure 6.6.17

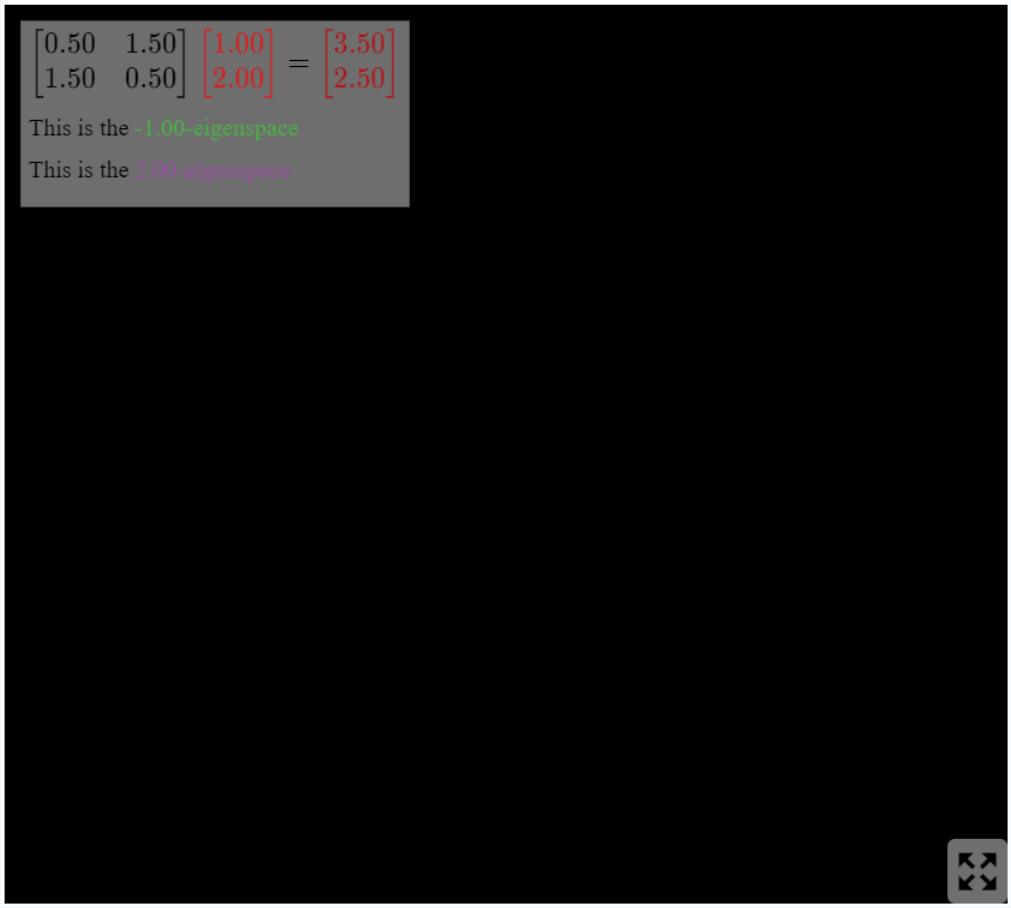


Figure 6.6.18: The eigenspaces of A are the lines through v_1 and v_2, These are the images under C of the coordinate axes, which are the eigenspaces of B.

Example 6.6.10: Interactive: Another matrix similar to B

Continuing with Example 6.6.6, let

$$A' = \frac{1}{5}\begin{pmatrix} -8 & -9 \\ 6 & 13 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \quad C' = \frac{1}{2}\begin{pmatrix} -1 & -3 \\ 2 & 1 \end{pmatrix},$$

so $A' = C'B(C')^{-1}$. Let $v_1' = \frac{1}{2}\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ and $v_2' = \frac{1}{2}\begin{pmatrix} -3 \\ 1 \end{pmatrix}$, the columns of C'. Then:

- B scales the e_1-direction by 2 and the e_2-direction by -1.
- A' scales the v_1'-direction by 2 and the v_2'-direction by -1.

As before, the x-axis is the 2-eigenspace of B, and the y-axis is the -1-eigenspace of B; likewise, the "v_1'-axis" is the 2-eigenspace of A', and the "v_2'-axis" is the -1-eigenspace of A'. This is consistent with fact 6.6.3, as multiplication by C' changes e_1 into $C'e_1 = v_1'$ and e_2 into $C'e_2 = v_2'$.

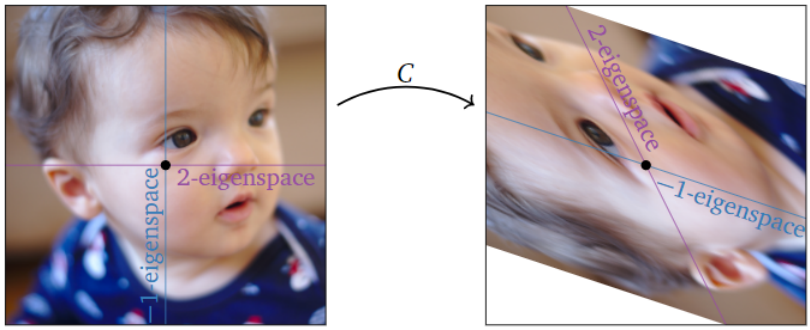


Figure 6.6.19

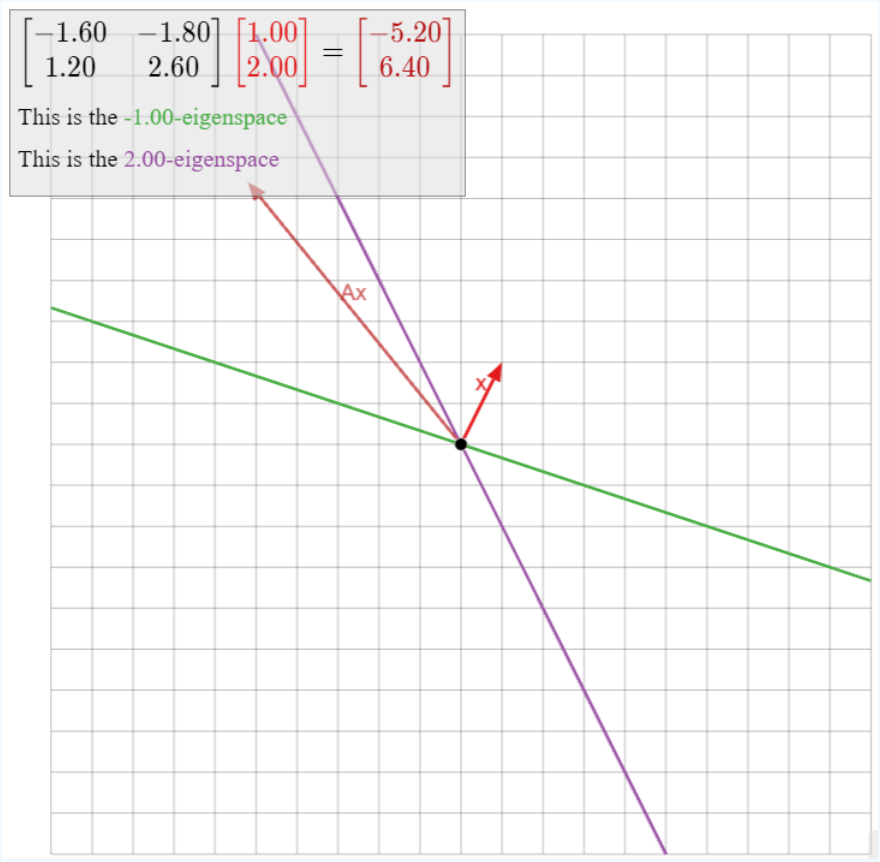


Figure 6.6.20: The eigenspaces of A' are the lines through v_1' and v_2'. These are the images under C' of the coordinate axes, which are the eigenspaces of B.

Example 6.6.11: Interactive: Similar 3×3 matrices

Continuing with Example 6.6.8, let

$$A = \begin{pmatrix} -1 & 0 & 0 \\ -1 & 0 & 2 \\ -1 & 1 & 1 \end{pmatrix} \quad B = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \quad C = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix},$$

so $A = CBC^{-1}$. Let v_1, v_2, v_3 by the columns of C. Then:

- B scales the e_1, e_2-plane by -1 and the e_3-direction by 2.
- A scales the v_1, v_2-plane by -1 and the v_3-direction by 2.

In other words, the xy-plane is the -1-eigenspace of B, and the z-axis is the 2-eigenspace of B; likewise, the "v_1, v_2-plane" is the -1-eigenspace of A, and the "v_3-axis" is the 2-eigenspace of A. This is consistent with fact 6.6.3, as multiplication by C changes e_1 into $Ce_1 = v_1$, e_2 into $Ce_2 = v_2$, and e_3 into $Ce_3 = v_3$.

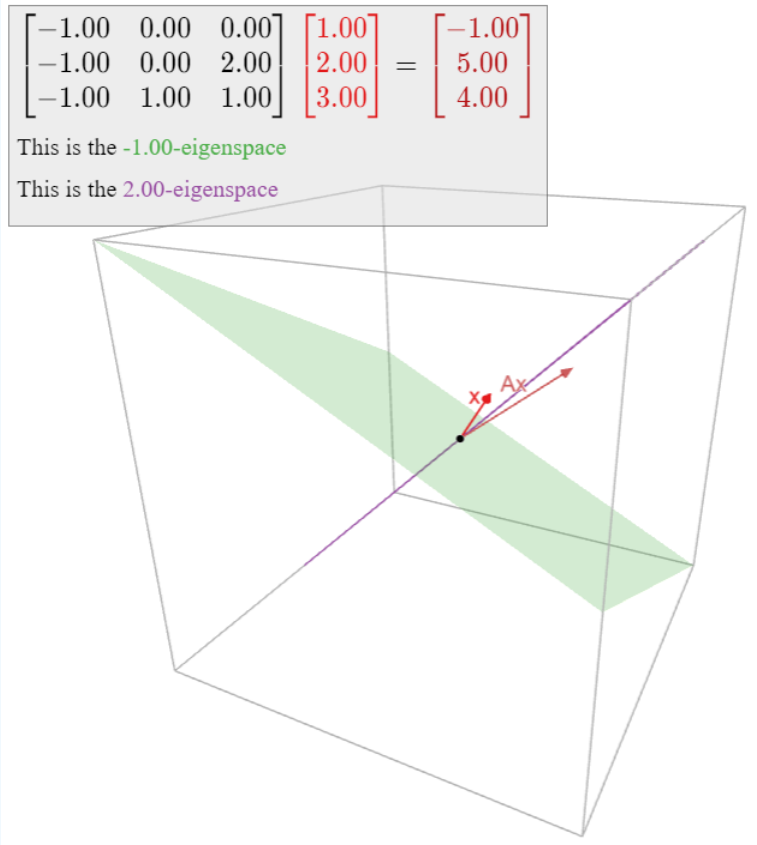


Figure 6.6.21: The -1-eigenspace of A is the green plane, and the 2-eigenspace of A is the violet line. These are the images under C of the xy-plane and the z-axis, respectively, which are the eigenspaces of B.

6.6: Similarity is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by LibreTexts.

- **5.3: Similarity** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

CHAPTER OVERVIEW

7: Orthogonality

7.1: Dot Products and Orthogonality
7.2: Orthogonal Complements
7.3: Orthogonal Projection
7.4: Orthogonal Sets and Gram-Schmidt Process
7.5: The Method of Least Squares
7.6: Orthogonal Diagonalization

This page titled 7: Orthogonality is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

7.1: Dot Products and Orthogonality

Learning Objectives

1. Understand the relationship between the dot product, length, and distance.
2. Understand the relationship between the dot product and orthogonality.
3. *Vocabulary words:* ***dot product, length, distance, unit vector, unit vector in the direction of*** x.
4. *Essential vocabulary word:* ***orthogonal***.

In this chapter, it will be necessary to find the *closest* point on a subspace to a given point, like so:

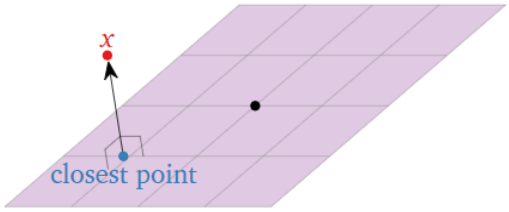


Figure 7.1.1

The closest point has the property that the difference between the two points is *orthogonal*, or *perpendicular*, to the subspace. For this reason, we need to develop notions of orthogonality, length, and distance.

The Dot Product

The basic construction in this section is the *dot product*, which measures angles between vectors and computes the length of a vector.

Definition 7.1.1: Dot Product

The ***dot product*** of two vectors x, y in $\mathbb{R}^n$ is

$$x \cdot y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

Thinking of x, y as column vectors, this is the same as $x^T y$.

For example,

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 32.$$

Notice that the dot product of two *vectors* is a *scalar*.

You can do arithmetic with dot products mostly as usual, as long as you remember you can only dot two vectors together, and that the result is a scalar.

Note 7.1.1: Properties of the Dot Product

Let x, y, z be vectors in $\mathbb{R}^n$ and let c be a scalar.

1. *Commutativity:* $x \cdot y = y \cdot x$.
2. *Distributivity with addition:* $(x + y) \cdot z = x \cdot z + y \cdot z$.
3. *Distributivity with scalar multiplication:* $(cx) \cdot y = c(x \cdot y)$.

The dot product of a vector with itself is an important special case:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1^2 + x_2^2 + \cdots + x_n^2.$$

Therefore, for any vector x, we have:

- $x \cdot x \geq 0$
- $x \cdot x = 0 \iff x = 0.$

This leads to a good definition of *length*.

Fact 7.1.1

The ***length*** of a vector x in $\mathbb{R}^n$ is the number

$$\|x\| = \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}\,.$$

It is easy to see why this is true for vectors in $\mathbb{R}^2$, by the Pythagorean theorem.

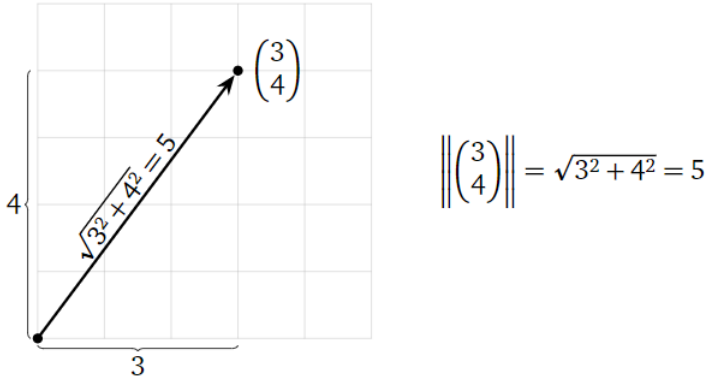


Figure 7.1.1

For vectors in $\mathbb{R}^3$, one can check that $\|x\|$ really is the length of x, although now this requires *two* applications of the Pythagorean theorem.

Note that the length of a vector is the length of the *arrow*; if we think in terms of points, then the length is its distance from the origin.

Example 7.1.1

Suppose that $\|x\| = 2,\ \|y\| = 3$, and $x \cdot y = -4$. What is $\|2x + 3y\|$?

Solution

We compute

$$\begin{aligned} \|2x+3y\|^2 &= (2x+3y)\cdot(2x+3y) \\ &= 4x\cdot x + 6x\cdot y + 6x\cdot y + 9y\cdot y \\ &= 4\|x\|^2 + 9\|y\|^2 + 12x\cdot y \\ &= 4\cdot 4 + 9\cdot 9 - 12\cdot 4 = 49. \end{aligned}$$

Hence $\|2x+3y\| = \sqrt{49} = 7$.

Fact 7.1.2

If x is a vector and c is a scalar, then $\|cx\| = |c| \cdot \|x\|$.

This says that scaling a vector by c scales its length by $|c|$. For example,

$$\left\| \begin{pmatrix} 6 \\ 8 \end{pmatrix} \right\| = \left\| 2 \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right\| = 2 \left\| \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right\| = 10.$$

Now that we have a good notion of length, we can define the *distance* between points in $\mathbb{R}^n$. Recall that the difference between two points x, y is naturally a vector, namely, the vector $y - x$ pointing from x to y.

Definition 7.1.2: Distance

The ***distance*** between two points x, y in $\mathbb{R}^n$ is the length of the vector from, Note 2.1.2 in Section 2.1, x to y:

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

Example 7.1.2

Find the distance from $(1, 2)$ to $(4, 4)$.

Solution

Let $x = (1, 2)$ and $y = (4, 4)$. Then

$$\text{dist}(x, y) = \|y - x\| = \left\| \begin{pmatrix} 3 \\ 2 \end{pmatrix} \right\| = \sqrt{3^2 + 2^2} = \sqrt{13}.$$

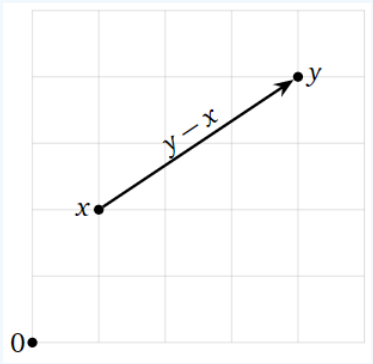


Figure 7.1.2

Vectors with length 1 are very common in applications, so we give them a name.

Definition 7.1.3: Unit Vector

A ***unit vector*** is a vector x with length $\|x\| = \sqrt{x \cdot x} = 1$.

The standard coordinate vectors, Note 3.3.2 in Section 3.3, $e_1, e_2, e_3, \ldots$ are unit vectors:

$$\|e_1\| = \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\| = \sqrt{1^2 + 0^2 + 0^2} = 1.$$

For any nonzero vector x, there is a unique unit vector pointing in the same direction. It is obtained by dividing by the length of x.

Fact 7.1.3

Let x be a nonzero vector in $\mathbb{R}^n$. The ***unit vector in the direction of x*** is the vector $x/\|x\|$.

This is in fact a unit vector (noting that $\|x\|$ is a positive number, so $|1/\|x\|| = 1/\|x\|$):

$$\left\| \frac{x}{\|x\|} \right\| = \frac{1}{\|x\|} \|x\| = 1.$$

Example 7.1.3

What is the unit vector u in the direction of $x = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$?

Solution

We divide by the length of x:

$$u = \frac{x}{\|x\|} = \frac{1}{\sqrt{3^2+4^2}} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 3/5 \\ 4/5 \end{pmatrix}.$$

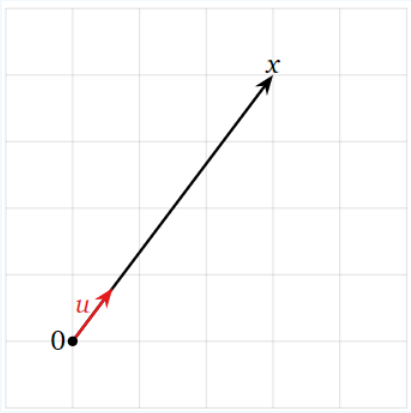


Figure 7.1.3

Orthogonal Vectors

In this section, we show how the dot product can be used to define *orthogonality*, i.e., when two vectors are perpendicular to each other.

Definition 7.1.4: Orthogonal and Perpendicular

Two vectors x, y in $\mathbb{R}^n$ are ***orthogonal*** or ***perpendicular*** if $x \cdot y = 0$.

Notation: $x \perp y$ means $x \cdot y = 0$.

Note 7.1.2

Since $0 \cdot x = 0$ for any vector x, the zero vector is orthogonal to every vector in $\mathbb{R}^n$.

We motivate the above definition using the *law of cosines* in $\mathbb{R}^2$. In our language, the law of cosines asserts that if x, y are two nonzero vectors, and if $\alpha > 0$ is the angle between them, then

$$\|y - x\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\| \, \|y\| \cos\alpha.$$

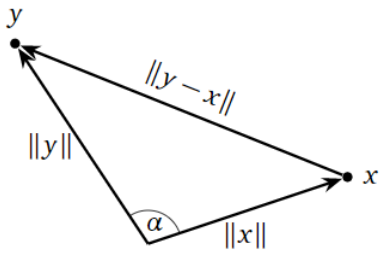


Figure 7.1.4

In particular, $\alpha = 90^\circ$ if and only if $\cos(\alpha) = 0$, which happens if and only if $\|y-x\|^2 = \|x\|^2 + \|y\|^2$. Therefore,

$$\begin{aligned} x \text{ and } y \text{ are perpendicular} &\iff \|x\|^2 + \|y\|^2 = \|y-x\|^2 \\ &\iff x\cdot x + y\cdot y = (y-x)\cdot(y-x) \\ &\iff x\cdot x + y\cdot y = y\cdot y + x\cdot x - 2x\cdot y \\ &\iff x\cdot y = 0. \end{aligned}$$

To reiterate:

Note 7.1.3

$$x \perp y \iff x\cdot y = 0 \iff \|y-x\|^2 = \|x\|^2 + \|y\|^2.$$

Example 7.1.4

Find all vectors orthogonal to $v = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$.

Solution

We have to find all vectors x such that $x \cdot v = 0$. This means solving the equation

$$0 = x\cdot v = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = x_1 + x_2 - x_3.$$

The parametric form for the solution set is $x_1 = -x_2 + x_3$, so the parametric vector form of the general solution is

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Therefore, the answer is the *plane*

$$\text{Span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

For instance,

$$\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \perp \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \quad \text{because} \quad \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = 0.$$

Example 7.1.5

Find all vectors orthogonal to both $v = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ and $w = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Solution

We have to solve the system of two homogeneous equations

$$\begin{aligned} 0 &= x \cdot v = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = x_1 + x_2 - x_3 \\ 0 &= x \cdot w = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = x_1 + x_2 + x_3. \end{aligned}$$

In matrix form:

$$\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The parametric vector form of the solution set is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

Therefore, the answer is the *line*

$$\text{Span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

For instance,

$$\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = 0 \quad \text{and} \quad \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0.$$

Remark: Angle between two vectors

More generally, the law of cosines gives a formula for the angle α between two nonzero vectors:

$$\begin{aligned} 2\|x\|\|y\|\cos(\alpha) &= \|x\|^2 + \|y\|^2 - \|y-x\|^2 \\ &= x \cdot x + y \cdot y - (y-x) \cdot (y-x) \\ &= x \cdot x + y \cdot y - y \cdot y - x \cdot x + 2x \cdot y \\ &= 2x \cdot y \\ \implies \alpha &= \cos^{-1}\left(\frac{x \cdot y}{\|x\|\|y\|} \right). \end{aligned}$$

In higher dimensions, we take this to be the *definition* of the angle between x and y.

This page titled 7.1: Dot Products and Orthogonality is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **6.1: Dot Products and Orthogonality** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

7.2: Orthogonal Complements

Learning Objectives

1. Understand the basic properties of orthogonal complements.
2. Learn to compute the orthogonal complement of a subspace.
3. *Recipes:* shortcuts for computing the orthogonal complements of common subspaces.
4. *Picture:* orthogonal complements in $\mathbb{R}^2$ and $\mathbb{R}^3$.
5. *Theorem:* row rank equals column rank.
6. *Vocabulary words:* ***orthogonal complement, row space***.

It will be important to compute the set of *all* vectors that are orthogonal to a given set of vectors. It turns out that a vector is orthogonal to a set of vectors if and only if it is orthogonal to the span of those vectors, which is a subspace, so we restrict ourselves to the case of subspaces.

Definition of the Orthogonal Complement

Taking the orthogonal complement is an operation that is performed on *subspaces*.

Definition 7.2.1: Orthogonal Complement

Let W be a subspace of $\mathbb{R}^n$. Its ***orthogonal complement*** is the subspace

$$W^\perp = \{v \text{ in } \mathbb{R}^n \mid v \cdot w = 0 \text{ for all } w \text{ in } W\}.$$

The symbol $W^\perp$ is sometimes read "W perp."

This is the set of all vectors v in $\mathbb{R}^n$ that are orthogonal to all of the vectors in W. We will show below 15 that $W^\perp$ is indeed a subspace.

Note 7.2.1

We now have two similar-looking pieces of notation:

$$A^T \text{ is the transpose of a matrix } A.$$
$$W^\perp \text{ is the orthogonal complement of a subspace } W.$$

Try not to confuse the two.

Pictures of orthogonal complements

The orthogonal complement of a line W through the origin in $\mathbb{R}^2$ is the perpendicular line $W^\perp$.

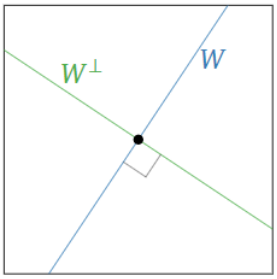


Figure 7.2.1

Example 7.2.1: Interactive: Orthogonal complements in $\mathbb{R}^2$

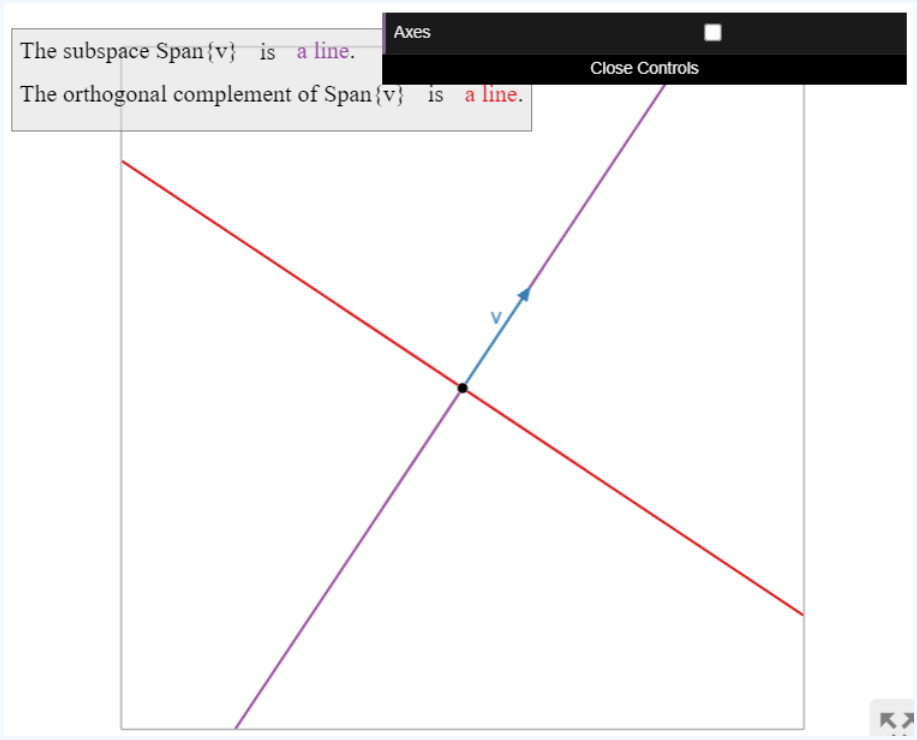


Figure 7.2.2: The orthogonal complement of the line spanned by v is the perpendicular line. Click and drag the head of v to move it.

The orthogonal complement of a line W in $\mathbb{R}^3$ is the perpendicular plane $W^\perp$.

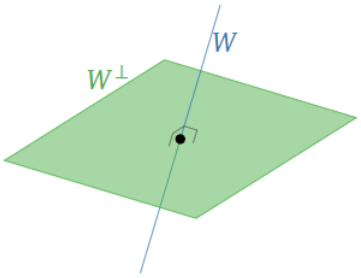


Figure 7.2.3

Example 7.2.2: Interactive: Orthogonal complements in $\mathbb{R}^3$

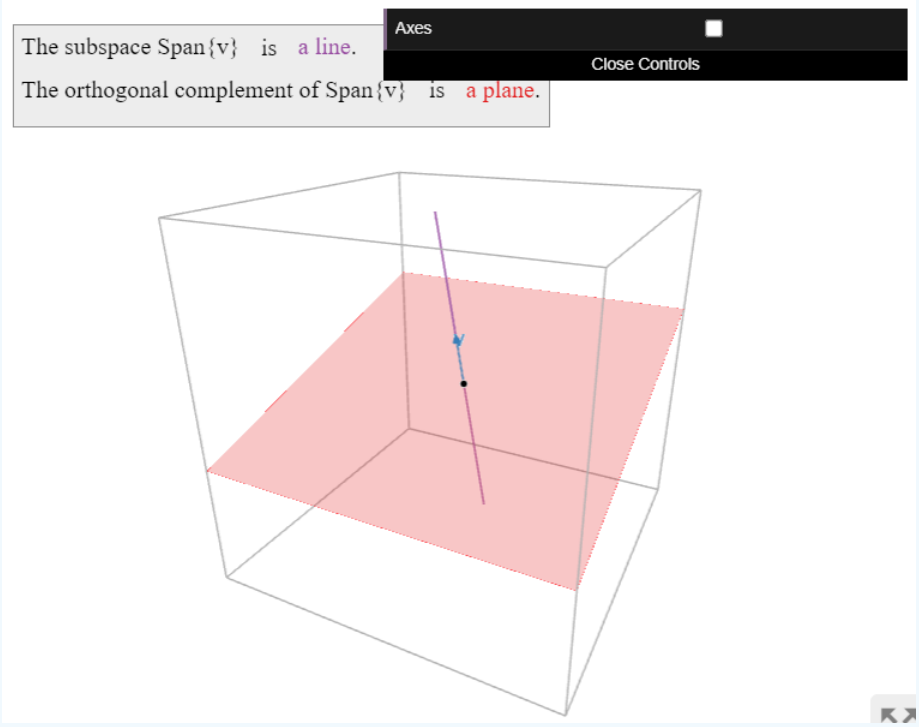


Figure 7.2.4: The orthogonal complement of the line spanned by v is the perpendicular plane. Click and drag the head of v to move it.

The orthogonal complement of a plane W in $\mathbb{R}^3$ is the perpendicular line $W^\perp$.

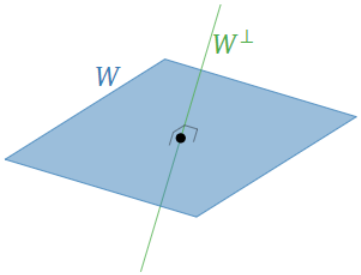


Figure 7.2.5

Example 7.2.3: Interactive: Orthogonal complements in $\mathbb{R}^3$

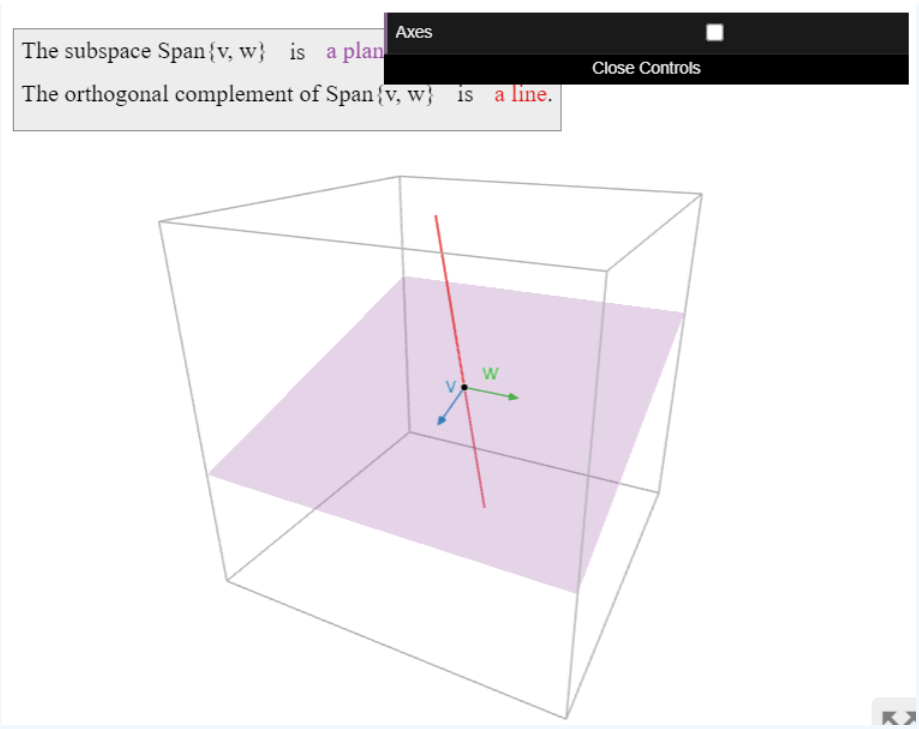


Figure 7.2.6: The orthogonal complement of the plane spanned by v, w is the perpendicular line. Click and drag the heads of v, w to change the plane.

We see in the above pictures that $(W^\perp)^\perp = W$.

Example 7.2.4

The orthogonal complement of $\mathbb{R}^n$ is $\{0\}$, since the zero vector is the only vector that is orthogonal to all of the vectors in $\mathbb{R}^n$.

For the same reason, we have $\{0\}^\perp = \mathbb{R}^n$.

Computing Orthogonal Complements

Since any subspace is a span, the following proposition gives a recipe for computing the orthogonal complement of any subspace. However, below we will give several shortcuts for computing the orthogonal complements of other common kinds of subspaces–in particular, null spaces. To compute the orthogonal complement of a general subspace, usually it is best to rewrite the subspace as the column space or null space of a matrix, as in Note 2.6.3 in Section 2.6.

Proposition 7.2.1: The Orthogonal Complement of a Column Space

Let A be a matrix and let $W = \text{Col}(A)$. Then

$$W^\perp = \text{Nul}(A^T).$$

Proof

To justify the first equality, we need to show that a vector x is perpendicular to the all of the vectors in W if and only if it is perpendicular only to $v_1, v_2, \ldots, v_m$. Since the v_i are contained in W, we really only have to show that if $x \cdot v_1 = x \cdot v_2 = \cdots = x \cdot v_m = 0$, then x is perpendicular to every vector v in W. Indeed, any vector in W has the form $v = c_1v_1 + c_2v_2 + \cdots + c_mv_m$ for suitable scalars $c_1, c_2, \ldots, c_m$, so

$$\begin{aligned} x \cdot v &= x \cdot (c_1v_1 + c_2v_2 + \cdots + c_mv_m) \\ &= c_1(x \cdot v_1) + c_2(x \cdot v_2) + \cdots + c_m(x \cdot v_m) \\ &= c_1(0) + c_2(0) + \cdots + c_m(0) = 0. \end{aligned}$$

Therefore, x is in $W^\perp$.

To prove the second equality, we let

$$A = \begin{pmatrix} -v_1^T- \\ -v_2^T- \\ \vdots \\ -v_m^T- \end{pmatrix}.$$

By the row-column rule for matrix multiplication Definition 2.3.3 in Section 2.3, for any vector x in $\mathbb{R}^n$ we have

$$Ax = \begin{pmatrix} v_1^T x \\ v_2^T x \\ \vdots \\ v_m^T x \end{pmatrix} = \begin{pmatrix} v_1 \cdot x \\ v_2 \cdot x \\ \vdots \\ v_m \cdot x \end{pmatrix}.$$

Therefore, x is in $\text{Nul}(A)$ if and only if x is perpendicular to each vector $v_1, v_2, \ldots, v_m$.

Since column spaces are the same as spans, we can rephrase the proposition as follows. Let $v_1, v_2, \ldots, v_m$ be vectors in $\mathbb{R}^n$, and let $W = \text{Span}\{v_1, v_2, \ldots, v_m\}$. Then

$$W^\perp = \{\text{all vectors orthogonal to each } v_1, v_2, \ldots, v_m\} = \text{Nul} \begin{pmatrix} -v_1^T- \\ -v_2^T- \\ \vdots \\ -v_m^T- \end{pmatrix}.$$

Again, it is important to be able to go easily back and forth between spans and column spaces. If you are handed a span, you can apply the proposition once you have rewritten your span as a column space.

By the proposition, computing the orthogonal complement of a span means *solving a system of linear equations*. For example, if

$$v_1 = \begin{pmatrix} 1 \\ 7 \\ 2 \end{pmatrix} \qquad v_2 = \begin{pmatrix} -2 \\ 3 \\ 1 \end{pmatrix}$$

then $\text{Span}\{v_1, v_2\}^\perp$ is the solution set of the homogeneous linear system associated to the matrix

$$\begin{pmatrix} -v_1^T- \\ -v_2^T- \end{pmatrix} = \begin{pmatrix} 1 & 7 & 2 \\ -2 & 3 & 1 \end{pmatrix}.$$

This is the solution set of the system of equations

$$\begin{cases} x_1 + 7x_2 + 2x_3 = 0 \\ -2x_1 + 3x_2 + x_3 = 0. \end{cases}$$

Example 7.2.5

Compute $W^\perp$, where

$$W = \text{Span}\left\{ \begin{pmatrix} 1 \\ 7 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 3 \\ 1 \end{pmatrix} \right\}.$$

Solution

According to Proposition 7.2.1, we need to compute the null space of the matrix

$$\begin{pmatrix} 1 & 7 & 2 \\ -2 & 3 & 1 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & -1/17 \\ 0 & 1 & 5/17 \end{pmatrix}.$$

The free variable is x_3, so the parametric form of the solution set is $x_1 = x_3/17$, $x_2 = -5x_3/17$, and the parametric vector form is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} 1/17 \\ -5/17 \\ 1 \end{pmatrix}.$$

Scaling by a factor of 17, we see that

$$W^\perp = \operatorname{Span}\left\{ \begin{pmatrix} 1 \\ -5 \\ 17 \end{pmatrix} \right\}.$$

We can check our work:

$$\begin{pmatrix} 1 \\ 7 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -5 \\ 17 \end{pmatrix} = 0 \qquad \begin{pmatrix} -2 \\ 3 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -5 \\ 17 \end{pmatrix} = 0.$$

Example 7.2.6

Find all vectors orthogonal to $v = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$.

Solution

According to Proposition 7.2.1, we need to compute the null space of the matrix

$$A = \begin{pmatrix} — & v & — \end{pmatrix} = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}.$$

This matrix is in reduced-row echelon form. The parametric form for the solution set is $x_1 = -x_2 + x_3$, so the parametric vector form of the general solution is

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Therefore, the answer is the *plane*

$$\operatorname{Span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

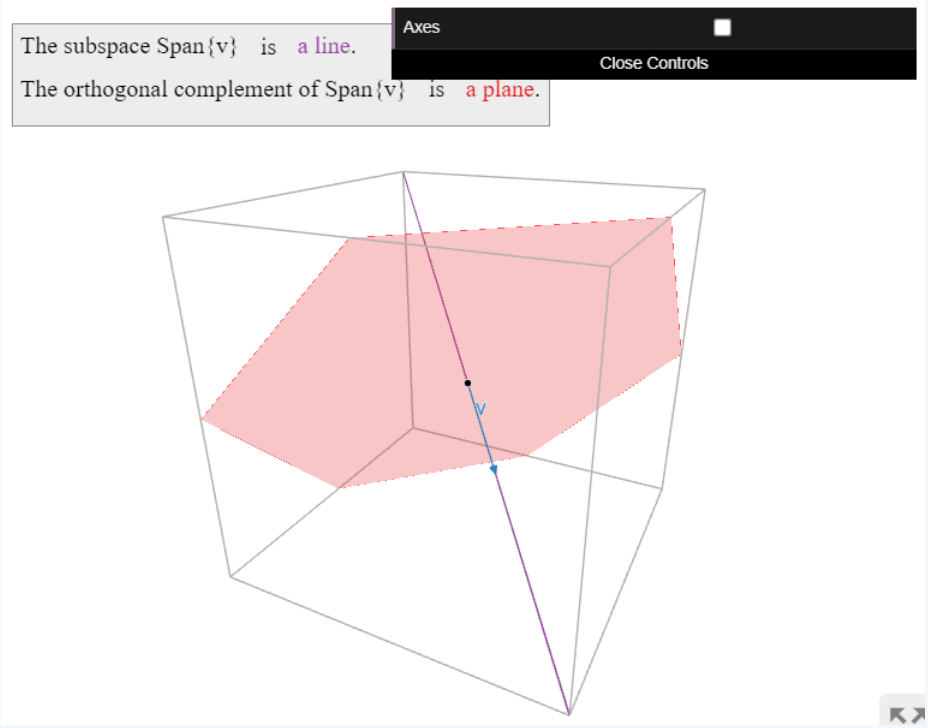


Figure 7.2.7: The set of all vectors perpendicular to v.

Example 7.2.7

Compute

$$\text{Span}\left\{\begin{pmatrix}1\\1\\-1\end{pmatrix},\begin{pmatrix}1\\1\\1\end{pmatrix}\right\}^{\perp}.$$

Solution

According to Proposition 7.2.1, we need to compute the null space of the matrix

$$A=\begin{pmatrix}1&1&-1\\1&1&1\end{pmatrix}\xrightarrow{\text{RREF}}\begin{pmatrix}1&1&0\\0&0&1\end{pmatrix}.$$

The parametric vector form of the solution is

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=x_2\begin{pmatrix}-1\\1\\0\end{pmatrix}.$$

Therefore, the answer is the *line*

$$\text{Span}\left\{\begin{pmatrix}-1\\1\\0\end{pmatrix}\right\}.$$

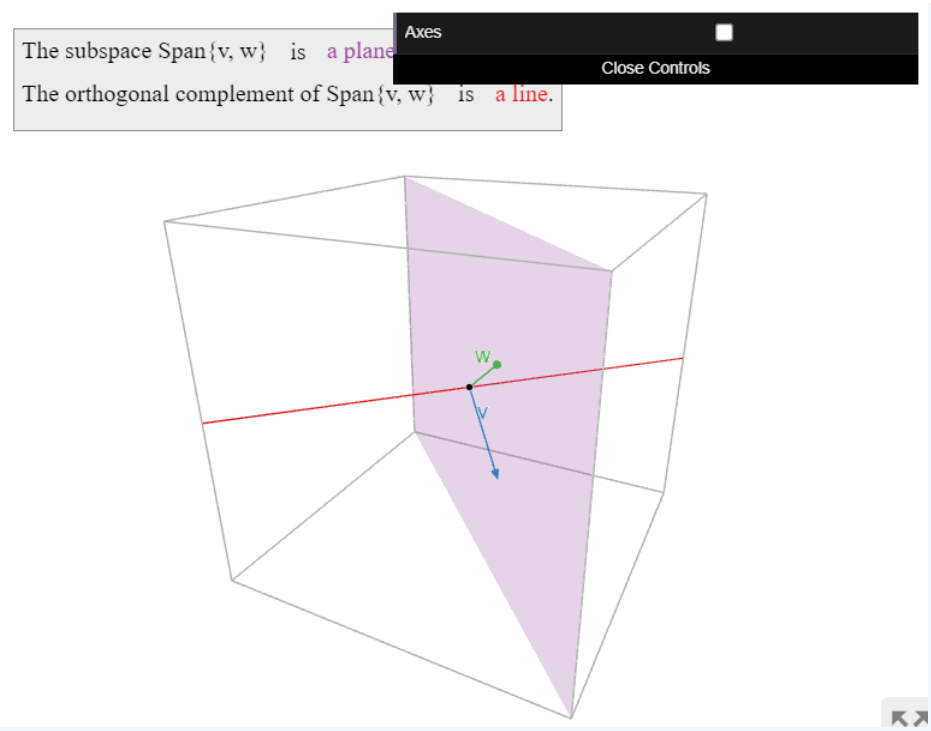


Figure 7.2.8: The orthogonal complement of the plane spanned by $v = (1, 1, -1)$ and $w = (1, 1, 1)$.

In order to find shortcuts for computing orthogonal complements, we need the following basic facts. Looking back the the above examples, all of these facts should be believable.

Fact 7.2.1: Facts about Orthogonal Complements

Let W be a subspace of $\mathbb{R}^n$. Then:

1. $W^\perp$ is also a subspace of $\mathbb{R}^n$.
2. $(W^\perp)^\perp = W$.
3. $\dim(W) + \dim(W^\perp) = n$.

Proof

For the first assertion, we verify the three defining properties of subspaces, Definition 2.6.2 in Section 2.6.

1. The zero vector is in $W^\perp$ because the zero vector is orthogonal to every vector in $\mathbb{R}^n$.
2. Let u, v be in $W^\perp$, so $u \cdot x = 0$ and $v \cdot x = 0$ for every vector x in W. We must verify that $(u + v) \cdot x = 0$ for every x in W. Indeed, we have
$$(u + v) \cdot x = u \cdot x + v \cdot x = 0 + 0 = 0.$$
3. Let u be in $W^\perp$, so $u \cdot x = 0$ for every x in W, and let c be a scalar. We must verify that $(cu) \cdot x = 0$ for every x in W. Indeed, we have
$$(cu) \cdot x = c(u \cdot x) = c0 = 0.$$

Next we prove the third assertion. Let $v_1, v_2, \ldots, v_m$ be a basis for W, so $m = \dim(W)$, and let $v_{m+1}, v_{m+2}, \ldots, v_k$ be a basis for $W^\perp$, so $k - m = \dim(W^\perp)$. We need to show $k = n$. First we claim that $\{v_1, v_2, \ldots, v_m, v_{m+1}, v_{m+2}, \ldots, v_k\}$ is linearly independent. Suppose that $c_1v_1 + c_2v_2 + \cdots + c_kv_k = 0$. Let $w = c_1v_1 + c_2v_2 + \cdots + c_mv_m$ and $w' = c_{m+1}v_{m+1} + c_{m+2}v_{m+2} + \cdots + c_kv_k$, so w is in W, w' is in W', and $w + w' = 0$. Then $w = -w'$ is in both W and $W^\perp$, which implies w is perpendicular to *itself*. In particular, $w \cdot w = 0$, so $w = 0$, and hence $w' = 0$. Therefore, all coefficients c_i are equal to zero, because $\{v_1, v_2, \ldots, v_m\}$ and $\{v_{m+1}, v_{m+2}, \ldots, v_k\}$ are linearly independent.

It follows from the previous paragraph that $k \leq n$. Suppose that $k < n$. Then the matrix

$$A = \begin{pmatrix} -v_1^T- \\ -v_2^T- \\ \vdots \\ -v_k^T- \end{pmatrix}$$

has more columns than rows (it is "wide"), so its null space is nonzero by Note 3.2.1 in Section 3.2. Let x be a nonzero vector in $\text{Nul}(A)$. Then

$$0 = Ax = \begin{pmatrix} v_1^T x \\ v_2^T x \\ \vdots \\ v_k^T x \end{pmatrix} = \begin{pmatrix} v_1 \cdot x \\ v_2 \cdot x \\ \vdots \\ v_k \cdot x \end{pmatrix}$$

by the row-column rule for matrix multiplication Definition 2.3.3 in Section 2.3. Since $v_1 \cdot x = v_2 \cdot x = \cdots = v_m \cdot x = 0$, it follows from Proposition 7.2.1 that x is in $W^\perp$, and similarly, x is in $(W^\perp)^\perp$. As above, this implies x is orthogonal to itself, which contradicts our assumption that x is nonzero. Therefore, $k = n$, as desired.

Finally, we prove the second assertion. Clearly W is contained in $(W^\perp)^\perp$: this says that everything in W is perpendicular to the set of all vectors perpendicular to everything in W. Let $m = \dim(W)$. By 3, we have $\dim(W^\perp) = n - m$, so $\dim((W^\perp)^\perp) = n - (n-m) = m$. The only m-dimensional subspace of $(W^\perp)^\perp$ is all of $(W^\perp)^\perp$, so $(W^\perp)^\perp = W$.

See subsection Pictures of orthogonal complements, for pictures of the second property. As for the third: for example, if W is a (2-dimensional) plane in $\mathbb{R}^4$, then $W^\perp$ is another (2-dimensional) plane. Explicitly, we have

$$\begin{aligned} \text{Span}\{e_1, e_2\}^\perp &= \left\{ \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \text{ in } \mathbb{R} \;\middle|\; \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = 0 \text{ and } \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = 0 \right\} \\ &= \left\{ \begin{pmatrix} 0 \\ 0 \\ z \\ w \end{pmatrix} \text{ in } \mathbb{R}^4 \right\} = \text{Span}\{e_3, e_4\} : \end{aligned}$$

the orthogonal complement of the xy-plane is the zw-plane.

Definition 7.2.2: Row Space

The ***row space*** of a matrix A is the span of the rows of A, and is denoted $\text{Row}(A)$.

If A is an $m \times n$ matrix, then the rows of A are vectors with n entries, so $\text{Row}(A)$ is a subspace of $\mathbb{R}^n$. Equivalently, since the rows of A are the columns of A^T, the row space of A is the column space of A^T:

$$\text{Row}(A) = \text{Col}(A^T).$$

We showed in the above Proposition 7.2.3 that if A has rows $v_1^T, v_2^T, \ldots, v_m^T$, then

$$\text{Row}(A)^\perp = \text{Span}\{v_1, v_2, \ldots, v_m\}^\perp = \text{Nul}(A).$$

Taking orthogonal complements of both sides and using the second fact 7.2.1 gives

$$\text{Row}(A) = \text{Nul}(A)^\perp.$$

Replacing A by A^T and remembering that $\text{Row}(A) = \text{Col}(A^T)$ gives

$$\text{Col}(A)^\perp = \text{Nul}(A^T) \quad \text{and} \quad \text{Col}(A) = \text{Nul}(A^T)^\perp.$$

To summarize:

Recipe: Shortcuts for Computing Orthogonal Complements

For any vectors $v_1, v_2, \ldots, v_m$, we have

$$\text{Span}\{v_1, v_2, \ldots, v_m\}^\perp = \text{Nul}\begin{pmatrix} -v_1^T- \\ -v_2^T- \\ \vdots \\ -v_m^T- \end{pmatrix}.$$

For any matrix A, we have

$$\begin{aligned} \text{Row}(A)^\perp &= \text{Nul}(A) & \text{Nul}(A)^\perp &= \text{Row}(A) \\ \text{Col}(A)^\perp &= \text{Nul}(A^T) & \text{Nul}(A^T)^\perp &= \text{Col}(A). \end{aligned}$$

As mentioned in the beginning of this subsection, in order to compute the orthogonal complement of a general subspace, usually it is best to rewrite the subspace as the column space or null space of a matrix.

Example 7.2.8: Orthogonal complement of a subspace

Compute the orthogonal complement of the subspace

$$W = \{(x, y, z) \text{ in } \mathbb{R}^3 \mid 3x + 2y = z\}.$$

Solution

Rewriting, we see that W is the solution set of the system of equations $3x + 2y - z = 0$, i.e., the null space of the matrix $A = \begin{pmatrix} 3 & 2 & -1 \end{pmatrix}$. Therefore,

$$W^\perp = \text{Row}(A) = \text{Span}\left\{\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}\right\}.$$

No row reduction was needed!

Example 7.2.9: Orthogonal complement of an eigenspace

Find the orthogonal complement of the 5-eigenspace of the matrix

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

Solution

The 5-eigenspace is

$$W = \text{Nul}(A - 5I_3) = \text{Nul}\begin{pmatrix} -3 & 4 & -1 \\ 3 & -3 & 0 \\ -2 & 4 & -2 \end{pmatrix},$$

so

$$W^\perp = \text{Row}\begin{pmatrix} -3 & 4 & -1 \\ 3 & -3 & 0 \\ -2 & 4 & -2 \end{pmatrix} = \text{Span}\left\{\begin{pmatrix} -3 \\ 4 \\ -1 \end{pmatrix}, \begin{pmatrix} 3 \\ -3 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 4 \\ -2 \end{pmatrix}\right\}.$$

These vectors are necessarily linearly dependent (why)?

Row rank and Column Rank

Suppose that A is an $m \times n$ matrix. Let us refer to the dimensions of $\mathrm{Col}(A)$ and $\mathrm{Row}(A)$ as the *row rank* and the *column rank* of A (note that the column rank of A is the same as the rank of A). The next theorem says that the row and column ranks are the same. This is surprising for a couple of reasons. First, $\mathrm{Row}(A)$ lies in $\mathbb{R}^n$ and $\mathrm{Col}(A)$ lies in $\mathbb{R}^m$. Also, the theorem implies that A and A^T have the same number of pivots, even though the reduced row echelon forms of A and A^T have nothing to do with each other otherwise.

Theorem 7.2.1

Let A be a matrix. Then the row rank of A is equal to the column rank of A.

Proof

By Theorem 2.9.1 in Section 2.9, we have

$$\dim \mathrm{Col}(A) + \dim \mathrm{Nul}(A) = n.$$

On the other hand the third fact $7.2.1$says that

$$\dim \mathrm{Nul}(A)^\perp + \dim \mathrm{Nul}(A) = n,$$

which implies $\dim \mathrm{Col}(A) = \dim \mathrm{Nul}(A)^\perp$. Since $\mathrm{Nul}(A)^\perp = \mathrm{Row}(A)$, we have

$$\dim \mathrm{Col}(A) = \dim \mathrm{Row}(A),$$

as desired.

In particular, by Corollary 2.7.1 in Section 2.7 both the row rank and the column rank are equal to the number of pivots of A.

This page titled 7.2: Orthogonal Complements is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **6.2: Orthogonal Complements** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

7.3: Orthogonal Projection

Learning Objectives

1. Understand the orthogonal decomposition of a vector with respect to a subspace.
2. Understand the relationship between orthogonal decomposition and orthogonal projection.
3. Understand the relationship between orthogonal decomposition and the closest vector on / distance to a subspace.
4. Learn the basic properties of orthogonal projections as linear transformations and as matrix transformations.
5. *Recipes:* orthogonal projection onto a line, orthogonal decomposition by solving a system of equations, orthogonal projection via a complicated matrix product.
6. *Pictures:* orthogonal decomposition, orthogonal projection.
7. *Vocabulary words:* ***orthogonal decomposition***, ***orthogonal projection***.

Let W be a subspace of $\mathbb{R}^n$ and let x be a vector in $\mathbb{R}^n$. In this section, we will learn to compute the *closest vector* x_W to x in W. The vector x_W is called the *orthogonal projection* of x onto W. This is exactly what we will use to almost solve matrix equations, as discussed in the introduction to Chapter 6.

Orthogonal Decomposition

We begin by fixing some notation.

Definition 7.3.1: Notation

Let W be a subspace of $\mathbb{R}^n$ and let x be a vector in $\mathbb{R}^n$. We denote the closest vector to x on W by x_W.

To say that x_W is the closest vector to x on W means that the difference $x - x_W$ is *orthogonal* to the vectors in W:

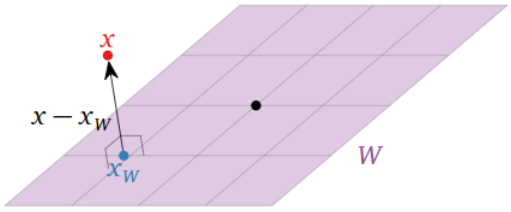


Figure 7.3.1

In other words, if $x_{W^\perp} = x - x_W$, then we have $x = x_W + x_{W^\perp}$, where x_W is in W and $x_{W^\perp}$ is in $W^\perp$. The first order of business is to prove that the closest vector always exists.

Theorem 7.3.1: Orthogonal Decomposition

Let W be a subspace of $\mathbb{R}^n$ and let x be a vector in $\mathbb{R}^n$. Then we can write x uniquely as

$$x = x_W + x_{W^\perp}$$

where x_W is the closest vector to x on W and $x_{W^\perp}$ is in $W^\perp$.

Proof

Let $m = \dim(W)$, so $n - m = \dim(W^\perp)$ by Fact 6.2.1 in Section 6.2. Let $v_1, v_2, \ldots, v_m$ be a basis for W and let $v_{m+1}, v_{m+2}, \ldots, v_n$ be a basis for $W^\perp$. We showed in the proof of this Fact 6.2.1 in Section 6.2 that $\{v_1, v_2, \ldots, v_m, v_{m+1}, v_{m+2}, \ldots, v_n\}$ is linearly independent, so it forms a basis for $\mathbb{R}^n$. Therefore, we can write

$$x = (c_1 v_1 + \cdots + c_m v_m) + (c_{m+1} v_{m+1} + \cdots + c_n v_n) = x_W + x_{W^\perp},$$

where $x_W = c_1 v_1 + \cdots + c_m v_m$ and $x_{W^\perp} = c_{m+1} v_{m+1} + \cdots + c_n v_n$. Since $x_{W^\perp}$ is orthogonal to W, the vector x_W is the closest vector to x on W, so this proves that such a decomposition exists.

As for uniqueness, suppose that

$$x = x_W + x_{W^\perp} = y_W + y_{W^\perp}$$

for x_W, y_W in W and $x_{W^\perp}, y_{W^\perp}$ in $W^\perp$. Rearranging gives

$$x_W - y_W = y_{W^\perp} - x_{W^\perp}.$$

Since W and $W^\perp$ are subspaces, the left side of the equation is in W and the right side is in $W^\perp$. Therefore, $x_W - y_W$ is in W and in $W^\perp$, so it is orthogonal to itself, which implies $x_W - y_W = 0$. Hence $x_W = y_W$ and $x_{W^\perp} = y_{W^\perp}$, which proves uniqueness.

Definition 7.3.2: Orthogonal Decomposition and Orthogonal Projection

Let W be a subspace of $\mathbb{R}^n$ and let x be a vector in $\mathbb{R}^n$. The expression

$$x = x_W + x_{W^\perp}$$

for x_W in W and $x_{W^\perp}$ in $W^\perp$, is called the ***orthogonal decomposition*** of x with respect to W, and the closest vector x_W is the ***orthogonal projection*** of x onto W.

Since x_W is the closest vector on W to x, the distance from x to the subspace W is the length of the vector from x_W to x, i.e., the length of $x_{W^\perp}$. To restate:

Note 7.3.1: Closest Vector and Distance

Let W be a subspace of $\mathbb{R}^n$ and let x be a vector in $\mathbb{R}^n$.

- The orthogonal projection x_W is the closest vector to x in W.
- The distance from x to W is $\|x_{W^\perp}\|$.

Example 7.3.1: Orthogonal decomposition with respect to the xy-plane

Let W be the xy-plane in $\mathbb{R}^3$, so $W^\perp$ is the z-axis. It is easy to compute the orthogonal decomposition of a vector with respect to this W:

$$x = \begin{pmatrix}1\\2\\3\end{pmatrix} \implies x_W = \begin{pmatrix}1\\2\\0\end{pmatrix} \quad x_{W^\perp} = \begin{pmatrix}0\\0\\3\end{pmatrix}$$
$$x = \begin{pmatrix}a\\b\\c\end{pmatrix} \implies x_W = \begin{pmatrix}a\\b\\0\end{pmatrix} \quad x_{W^\perp} = \begin{pmatrix}0\\0\\c\end{pmatrix}.$$

We see that the orthogonal decomposition in this case expresses a vector in terms of a "horizontal" component (in the xy-plane) and a "vertical" component (on the z-axis).

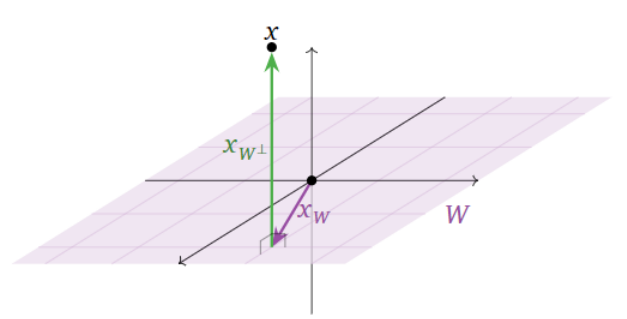


Figure 7.3.2

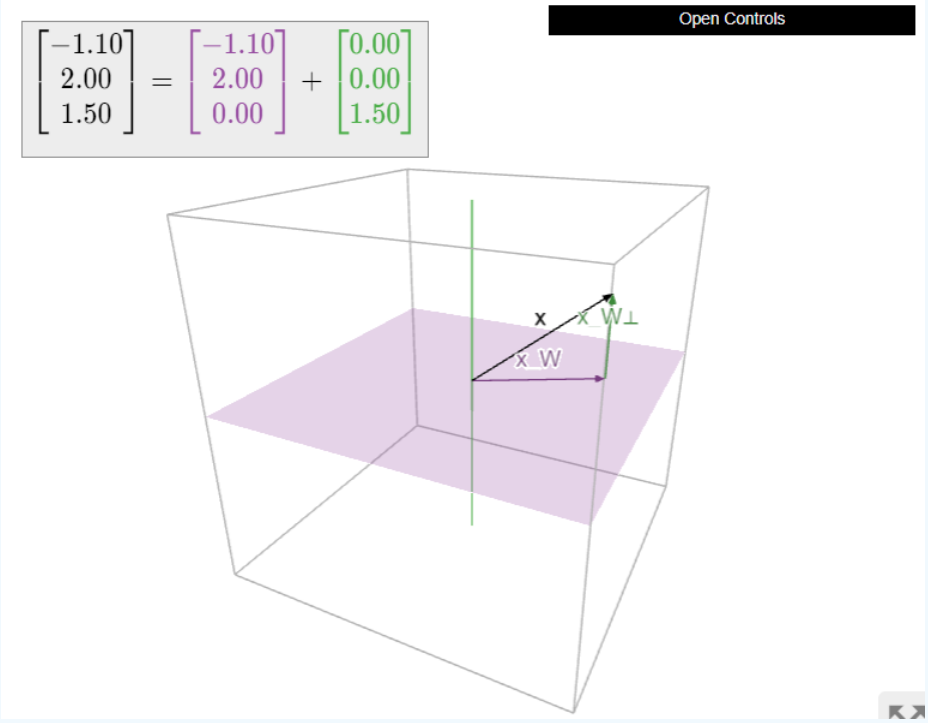


Figure 7.3.3: Orthogonal decomposition of a vector with respect to the xy-plane in $\mathbb{R}^3$. Note that x_W is in the xy-plane and $x_{W^\perp}$ is in the z-axis. Click and drag the head of the vector x to see how the orthogonal decomposition changes.

Example 7.3.2: Orthogonal decomposition of a vector in W

If x is in a subspace W, then the closest vector to x in W is itself, so $x = x_W$ and $x_{W^\perp} = 0$. Conversely, if $x = x_W$ then x is contained in W because x_W is contained in W.

Example 7.3.3: Orthogonal decomposition of a vector in $W^\perp$

If W is a subspace and x is in $W^\perp$, then the orthogonal decomposition of x is $x = 0 + x$, where 0 is in W and x is in $W^\perp$. It follows that $x_W = 0$. Conversely, if $x_W = 0$ then the orthogonal decomposition of x is $x = x_W + x_{W^\perp} = 0 + x_{W^\perp}$, so $x = x_{W^\perp}$ is in $W^\perp$.

Example 7.3.4: Interactive: Orthogonal decomposition in $\mathbb{R}^2$

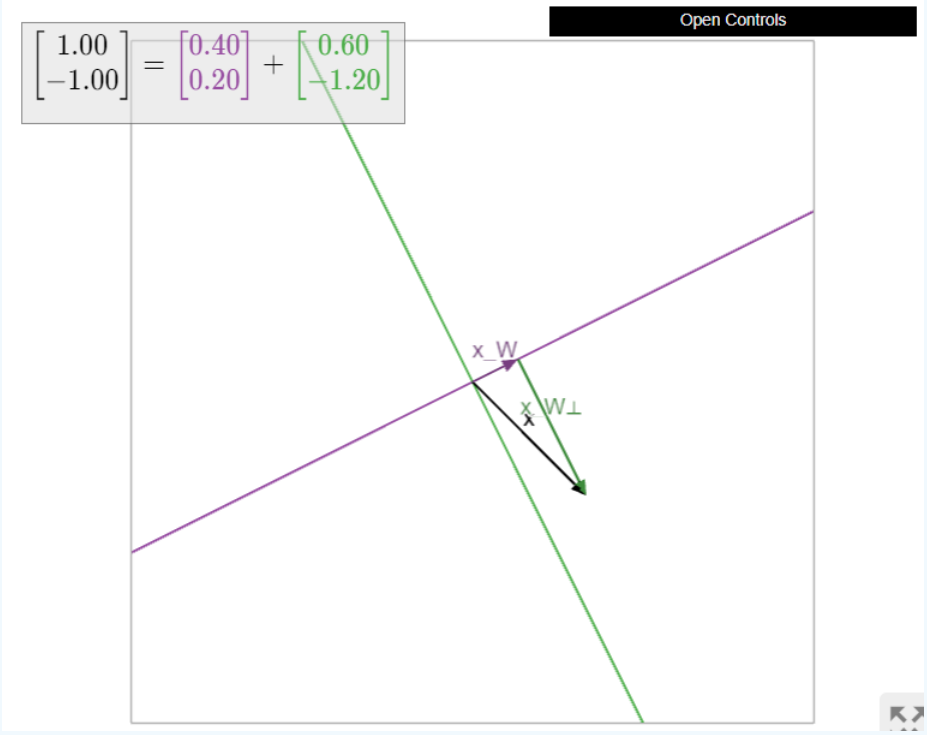


Figure 7.3.4: Orthogonal decomposition of a vector with respect to a line W in $\mathbb{R}^2$. Note that x_W is in W and $x_{W^\perp}$ is in the line perpendicular to W. Click and drag the head of the vector x to see how the orthogonal decomposition changes.

Example 7.3.5: Interactive: Orthogonal decomposition in $\mathbb{R}^3$

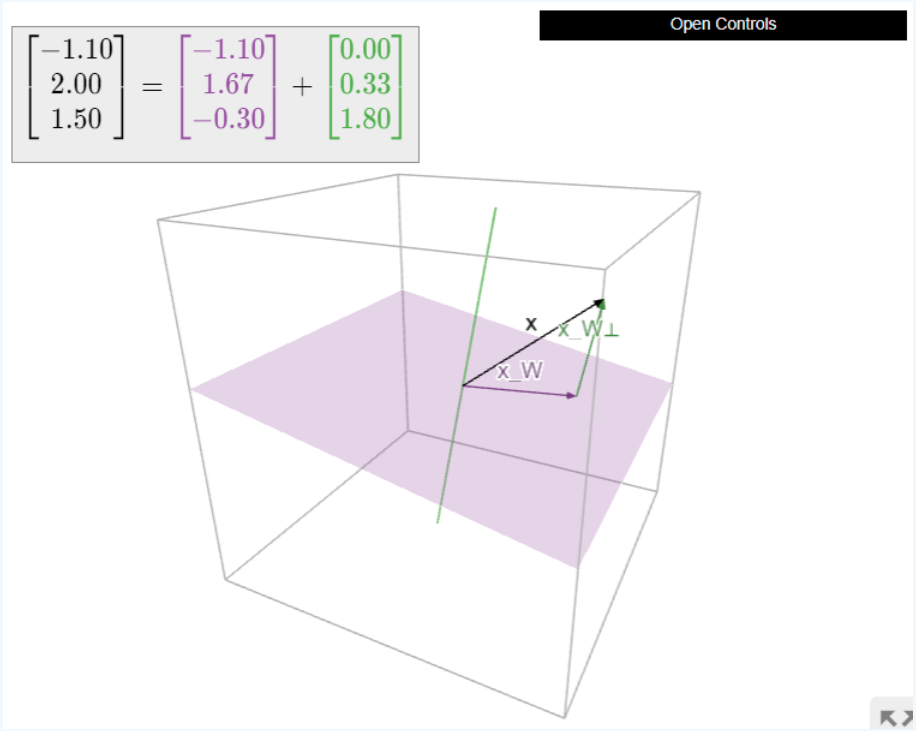


Figure 7.3.5: Orthogonal decomposition of a vector with respect to a plane W in $\mathbb{R}^3$. Note that x_W is in W and $x_{W^\perp}$ is in the line perpendicular to W. Click and drag the head of the vector x to see how the orthogonal decomposition changes.

Example 7.3.6: Interactive: Orthogonal decomposition in $\mathbb{R}^3$

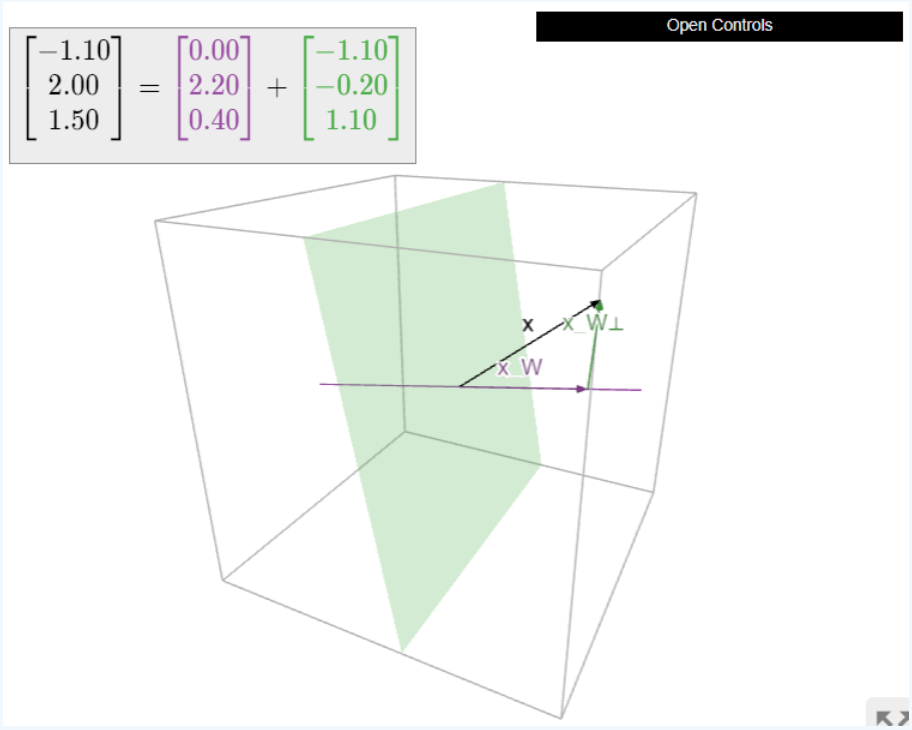


Figure 7.3.6: Orthogonal decomposition of a vector with respect to a line W in $\mathbb{R}^3$. Note that x_W is in W and $x_{W^\perp}$ is in the plane perpendicular to W. Click and drag the head of the vector x to see how the orthogonal decomposition changes.

Now we turn to the problem of computing x_W and $x_{W^\perp}$. Of course, since $x_{W^\perp} = x - x_W$, really all we need is to compute x_W. The following theorem gives a method for computing the orthogonal projection onto a column space. To compute the orthogonal projection onto a general subspace, usually it is best to rewrite the subspace as the column space of a matrix, as in Note 2.6.3 in Section 2.6.

Theorem7.3.2

Let A be an $m \times n$ matrix, let $W = \text{Col}(A)$, and let x be a vector in $\mathbb{R}^m$. Then the matrix equation

$$A^T Ac = A^T x$$

in the unknown vector c is consistent, and x_W is equal to Ac for any solution c.

Proof

Let $x = x_W + x_{W^\perp}$ be the orthogonal decomposition with respect to W. By definition x_W lies in $W = \text{Col}(A)$ and so there is a vector c in $\mathbb{R}^n$ with $Ac = x_W$. Choose any such vector c. We know that $x - x_W = x - Ac$ lies in $W^\perp$, which is equal to $\text{Nul}(A^T)$ by Recipe: Shortcuts for Computing Orthogonal Complements in Section 6.2. We thus have

$$0 = A^T(x - Ac) = A^T x - A^T Ac$$

and so

$$A^T Ac = A^T x.$$

This exactly means that $A^T Ac = A^T x$ is consistent. If c is any solution to $A^T Ac = A^T x$ then by reversing the above logic, we conclude that $x_W = Ac$.

Example 7.3.7: Orthogonal Projection onto a Line

Let $L = \text{Span}\{u\}$ be a line in $\mathbb{R}^n$ and let x be a vector in $\mathbb{R}^n$. By Theorem 7.3.2, to find x_L we must solve the matrix equation $u^T uc = u^T x$, where we regard u as an $n \times 1$ matrix (the column space of this matrix is exactly L!). But $u^T u = u \cdot u$ and $u^T x = u \cdot x$, so $c = (u \cdot x)/(u \cdot u)$ is a solution of $u^T uc = u^T x$, and hence $x_L = uc = (u \cdot x)/(u \cdot u)\, u$.

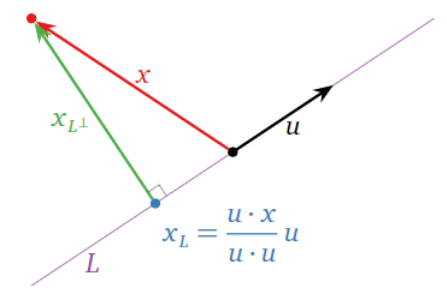


Figure 7.3.7

To reiterate:

Recipe: Orthogonal Projection onto a Line

If $L = \text{Span}\{u\}$ is a line, then

$$x_L = \frac{u \cdot x}{u \cdot u} u \quad \text{and} \quad x_{L^\perp} = x - x_L$$

for any vector x.

Remark: Simple proof for the formula for projection onto a line

In the special case where we are projecting a vector x in $\mathbb{R}^n$ onto a line $L = \text{Span}\{u\}$, our formula for the projection can be derived very directly and simply. The vector x_L is a multiple of u, say $x_L = cu$. This multiple is chosen so that $x - x_L = x - cu$ is perpendicular to u, as in the following picture.

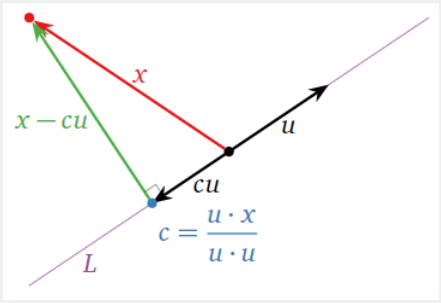


Figure 7.3.8

In other words,

$$(x - cu) \cdot u = 0.$$

Using the distributive property for the dot product and isolating the variable c gives us that

$$c = \frac{u \cdot x}{u \cdot u}$$

and so

$$x_L = cu = \frac{u \cdot x}{u \cdot u} u.$$

Example 7.3.8: Projection onto a line in $\mathbb{R}^2$

Compute the orthogonal projection of $x = \begin{pmatrix} -6 \\ 4 \end{pmatrix}$ onto the line L spanned by $u = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$, and find the distance from x to L.

Solution

First we find

$$x_L = \frac{x \cdot u}{u \cdot u} u = \frac{-18+8}{9+4} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = -\frac{10}{13} \begin{pmatrix} 3 \\ 2 \end{pmatrix} \qquad x_{L^\perp} = x - x_L = \frac{1}{13} \begin{pmatrix} -48 \\ 72 \end{pmatrix}.$$

The distance from x to L is

$$\|x_{L^\perp}\| = \frac{1}{13}\sqrt{48^2+72^2} \approx 6.656.$$

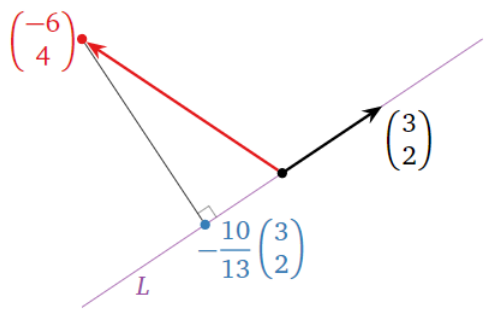


Figure 7.3.9

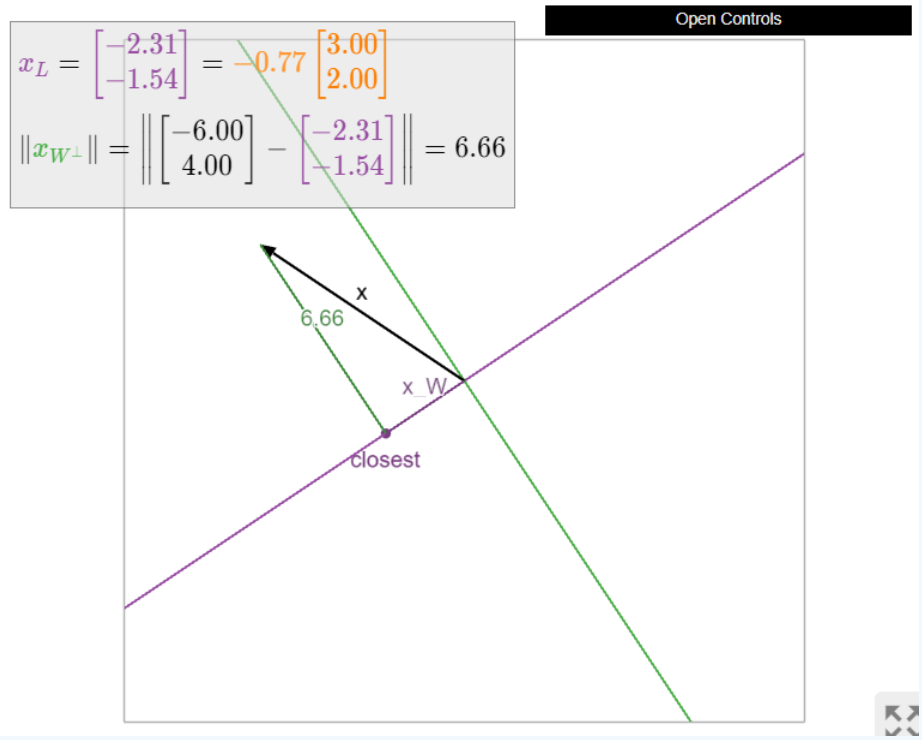


Figure 7.3.10: Distance from the line L.

Example 7.3.9: Projection onto a line in $\mathbb{R}^3$

Let

$$x = \begin{pmatrix} -2 \\ 3 \\ -1 \end{pmatrix} \qquad u = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix},$$

and let L be the line spanned by u. Compute x_L and $x_L^\perp$.

Solution

$$x_L = \frac{x \cdot u}{u \cdot u}\, u = \frac{2+3-1}{1+1+1} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} = \frac{4}{3} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \qquad x_{L^\perp} = x - x_L = \frac{1}{3} \begin{pmatrix} -2 \\ 5 \\ -7 \end{pmatrix}.$$

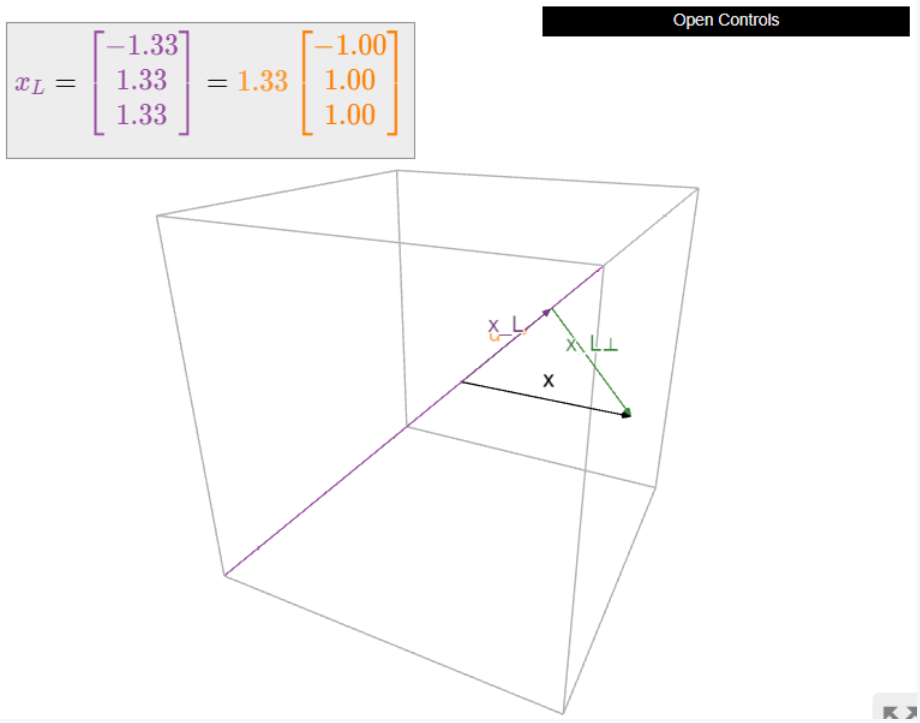


Figure 7.3.11: Orthogonal projection onto the line L.

When A is a matrix with more than one column, computing the orthogonal projection of x onto $W = \text{Col}(A)$ means solving the matrix equation $A^T Ac = A^T x$. In other words, we can compute the closest vector by *solving a system of linear equations*. To be explicit, we state Theorem 7.3.2 as a recipe:

Recipe: Compute an Orthogonal Decomposition

Let W be a subspace of $\mathbb{R}^m$. Here is a method to compute the orthogonal decomposition of a vector x with respect to W:

0. Rewrite W as the column space of a matrix A. In other words, find a a spanning set for W, and let A be the matrix with those columns.
1. Compute the matrix $A^T A$ and the vector $A^T x$.
2. Form the augmented matrix for the matrix equation $A^T Ac = A^T x$ in the unknown vector c, and row reduce.
3. This equation is always consistent; choose one solution c. Then

$$x_W = Ac \qquad x_{W^\perp} = x - x_W.$$

Example 7.3.10: Projection onto the xy-plane

Use Theorem 7.3.2 to compute the orthogonal decomposition of a vector with respect to the xy-plane in $\mathbb{R}^3$.

Solution

A basis for the xy-plane is given by the two standard coordinate vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Let A be the matrix with columns e_1, e_2:

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

Then

$$A^TA = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2 \qquad A^T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

It follows that the unique solution c of $A^TAc = I_2c = A^Tx$ is given by the first two coordinates of x, so

$$x_W = A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} \qquad x_{W^\perp} = x - x_W = \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix}.$$

We have recovered this Example 7.3.1.

Example 7.3.11: Projection onto a plane in $\mathbb{R}^3$

Let

$$W = \text{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\} \qquad x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Compute x_W and the distance from x to W.

Solution

We have to solve the matrix equation $A^TAc = A^Tx$, where

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

We have

$$A^TA = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \qquad A^Tx = \begin{pmatrix} -2 \\ 3 \end{pmatrix}.$$

We form an augmented matrix and row reduce:

$$\left(\begin{array}{cc|c} 2 & 1 & -2 \\ 1 & 2 & 3 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & -7/3 \\ 0 & 1 & 8/3 \end{array}\right) \implies c = \frac{1}{3}\begin{pmatrix} -7 \\ 8 \end{pmatrix}.$$

It follows that

$$x_W = Ac = \frac{1}{3}\begin{pmatrix}1\\8\\7\end{pmatrix} \qquad x_{W^\perp} = x - x_W = \frac{1}{3}\begin{pmatrix}2\\-2\\2\end{pmatrix}.$$

The distance from x to W is

$$\|x_{W^\perp}\| = \frac{1}{3}\sqrt{4+4+4} \approx 1.155.$$

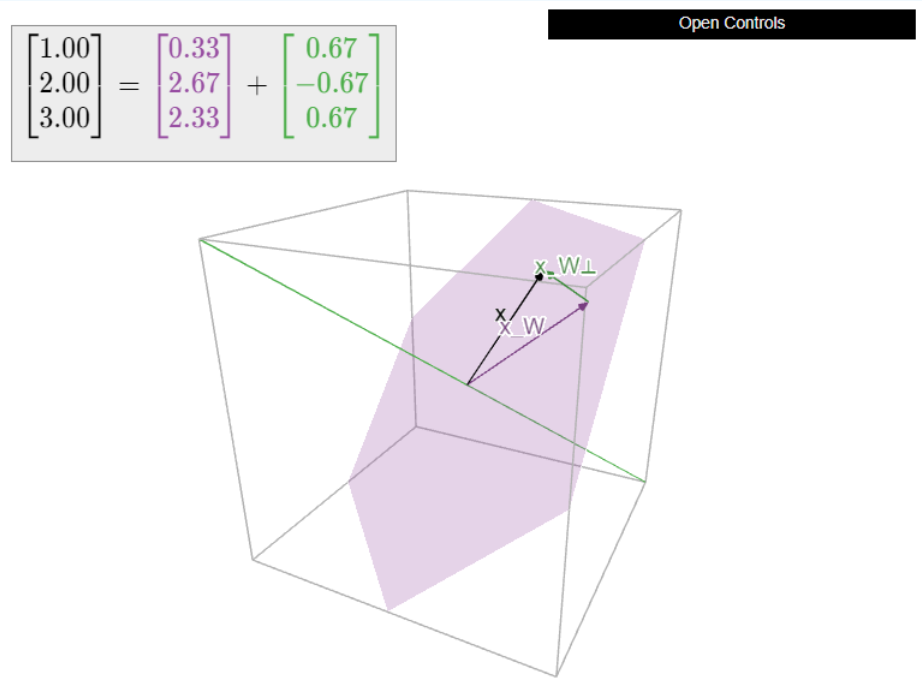


Figure 7.3.12: Orthogonal projection onto the plane W.

Example 7.3.12: Projection onto another plane in $\mathbb{R}^3$

Let

$$W = \left\{ \begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} \;\middle|\; x_1 - 2x_2 = x_3 \right\} \quad \text{and} \quad x = \begin{pmatrix}1\\1\\1\end{pmatrix}.$$

Compute x_W.

Solution

Method 1: First we need to find a spanning set for W. We notice that W is the solution set of the homogeneous equation $x_1 - 2x_2 - x_3 = 0$, so $W = \text{Nul}\begin{pmatrix}1 & -2 & -1\end{pmatrix}$. We know how to compute a basis for a null space: we row reduce and find the parametric vector form. The matrix $\begin{pmatrix}1 & -2 & -1\end{pmatrix}$ is already in reduced row echelon form. The parametric form is $x_1 = 2x_2 + x_3$, so the parametric vector form is

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = x_2\begin{pmatrix}2\\1\\0\end{pmatrix} + x_3\begin{pmatrix}1\\0\\1\end{pmatrix},$$

and hence a basis for V is given by

$$\left\{ \begin{pmatrix}2\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix} \right\}.$$

We let A be the matrix whose columns are our basis vectors:

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

Hence $\mathrm{Col}(A) = \mathrm{Nul}\begin{pmatrix} 1 & -2 & -1 \end{pmatrix} = W$.

Now we can continue with step 1 of the recipe. We compute

$$A^TA = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix} \qquad A^Tx = \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

We write the linear system $A^TAc = A^Tx$ as an augmented matrix and row reduce:

$$\left(\begin{array}{cc|c} 5 & 2 & 3 \\ 2 & 2 & 2 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & 1/3 \\ 0 & 1 & 2/3 \end{array}\right).$$

Hence we can take $c = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}$, so

$$x_W = Ac = \begin{pmatrix} 2 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 4 \\ 1 \\ 2 \end{pmatrix}.$$

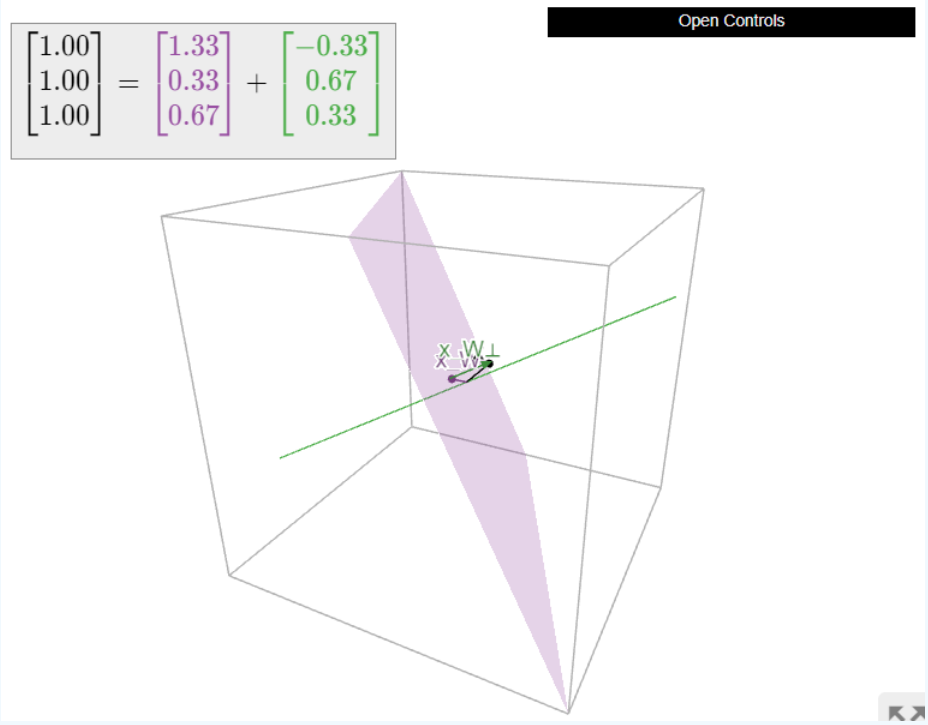


Figure 7.3.13: Orthogonal projection onto the plane W.

Method 2: In this case, it is easier to compute $x_{W^\perp}$. Indeed, since $W = \mathrm{Nul}\begin{pmatrix} 1 & -2 & -1 \end{pmatrix}$, the orthogonal complement is the line

$$V = W^\perp = \mathrm{Col}\begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}.$$

Using the formula for projection onto a line, Example 7.3.7, gives

$$x_{W^\perp} = x_V = \frac{\begin{pmatrix}1\\1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\-2\\-1\end{pmatrix}}{\begin{pmatrix}1\\-2\\-1\end{pmatrix}\cdot\begin{pmatrix}1\\-2\\-1\end{pmatrix}}\begin{pmatrix}1\\-2\\-1\end{pmatrix} = \frac{1}{3}\begin{pmatrix}-1\\2\\1\end{pmatrix}.$$

Hence we have

$$x_W = x - x_{W^\perp} = \begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{1}{3}\begin{pmatrix}-1\\2\\1\end{pmatrix}. = \frac{1}{3}\begin{pmatrix}4\\1\\2\end{pmatrix},$$

as above.

Example 7.3.13: Projection onto a 3-space in $\mathbb{R}^4$

Let

$$W = \text{Span}\left\{\begin{pmatrix}1\\0\\-1\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\\-1\end{pmatrix}, \begin{pmatrix}1\\1\\1\\-1\end{pmatrix}\right\} \qquad x = \begin{pmatrix}0\\1\\3\\4\end{pmatrix}.$$

Compute the orthogonal decomposition of x with respect to W.

Solution

We have to solve the matrix equation $A^TAc = A^Tx$, where

$$A = \begin{pmatrix}1&0&1\\0&1&1\\-1&0&1\\0&-1&-1\end{pmatrix}.$$

We compute

$$A^TA = \begin{pmatrix}2&0&0\\0&2&2\\0&2&4\end{pmatrix} \qquad A^Tx = \begin{pmatrix}-3\\3\\0\end{pmatrix}.$$

We form an augmented matrix and row reduce:

$$\left(\begin{array}{ccc|c}2&0&0&-3\\0&2&2&-3\\0&2&4&0\end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c}1&0&0&-3/2\\0&1&0&-3\\0&0&1&3/2\end{array}\right) \implies c = \frac{1}{2}\begin{pmatrix}-3\\-6\\3\end{pmatrix}.$$

It follows that

$$x_W = Ac = \frac{1}{2}\begin{pmatrix}0\\-3\\6\\3\end{pmatrix} \qquad x_{W^\perp} = \frac{1}{2}\begin{pmatrix}0\\5\\0\\5\end{pmatrix}.$$

In the context of the above recipe, if we start with a *basis* of W, then it turns out that the square matrix A^TA is automatically invertible! (It is always the case that A^TA is square and the equation $A^TAc = A^Tx$ is consistent, but A^TA need not be invertible in general.)

Corollary 7.3.1

Let A be an $m \times n$ matrix with linearly independent columns and let $W = \text{Col}(A)$. Then the $n \times n$ matrix A^TA is invertible, and for all vectors x in $\mathbb{R}^m$, we have

$$x_W = A(A^TA)^{-1}A^Tx.$$

Proof

We will show that $\text{Nul}(A^TA) = \{0\}$, which implies invertibility by Theorem 5.1.1 in Section 5.1. Suppose that $A^TAc = 0$. Then $A^TAc = A^T0$, so $0_W = Ac$ by Theorem 7.3.2 But $0_W = 0$ (the orthogonal decomposition of the zero vector is just $0 = 0 + 0$), so $Ac = 0$, and therefore c is in $\text{Nul}(A)$. Since the columns of A are linearly independent, we have $c = 0$, so $\text{Nul}(A^TA) = 0$, as desired.

Let x be a vector in $\mathbb{R}^n$ and let c be a solution of $A^TAc = A^Tx$. Then $c = (A^TA)^{-1}A^Tx$, so $x_W = Ac = A(A^TA)^{-1}A^Tx$.

The corollary applies in particular to the case where we have a subspace W of $\mathbb{R}^m$, and a basis $v_1, v_2, \ldots, v_n$ for W. To apply the corollary, we take A to be the $m \times n$ matrix with columns $v_1, v_2, \ldots, v_n$.

Example 7.3.14: Computing a projection

Continuing with the above Example 7.3.11, let

$$W = \text{Span}\left\{\begin{pmatrix}1\\0\\-1\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}\right\} \qquad x = \begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}.$$

Compute x_W using the formula $x_W = A(A^TA)^{-1}A^Tx$.

Solution

Clearly the spanning vectors are noncollinear, so according to Corollary 7.3.1, we have $x_W = A(A^TA)^{-1}A^Tx$, where

$$A = \begin{pmatrix}1&1\\0&1\\-1&0\end{pmatrix}.$$

We compute

$$A^TA = \begin{pmatrix}2&1\\1&2\end{pmatrix} \implies (A^TA)^{-1} = \frac{1}{3}\begin{pmatrix}2&-1\\-1&2\end{pmatrix},$$

so

$$\begin{aligned} x_W = A(A^TA)^{-1}A^Tx &= \begin{pmatrix}1&1\\0&1\\-1&0\end{pmatrix}\frac{1}{3}\begin{pmatrix}2&-1\\-1&2\end{pmatrix}\begin{pmatrix}1&0&-1\\1&1&0\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} \\ &= \frac{1}{3}\begin{pmatrix}2&1&-1\\1&2&1\\-1&1&2\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \frac{1}{3}\begin{pmatrix}2x_1 + x_2 - x_3\\x_1 + 2x_2 + x_3\\-x_1 + x_2 + 2x_3\end{pmatrix}. \end{aligned}$$

So, for example, if $x = (1, 0, 0)$, this formula tells us that $x_W = (2, 1, -1)$.

Orthogonal Projection

In this subsection, we change perspective and think of the orthogonal projection x_W as a *function* of x. This function turns out to be a linear transformation with many nice properties, and is a good example of a linear transformation which is not originally defined as a matrix transformation.

Proposition 7.3.1: Properties of Orthogonal Projections

Let W be a subspace of $\mathbb{R}^n$, and define $T\colon \mathbb{R}^n \to \mathbb{R}^n$ by $T(x) = x_W$. Then:

1. T is a linear transformation.
2. $T(x) = x$ if and only if x is in W.
3. $T(x) = 0$ if and only if x is in $W^\perp$.
4. $T \circ T = T$.
5. The range of T is W.

Proof

1. We have to verify the defining properties of linearity, Definition 3.3.1 in Section 3.3. Let x, y be vectors in $\mathbb{R}^n$, and let $x = x_W + x_{W^\perp}$ and $y = y_W + y_{W^\perp}$ be their orthogonal decompositions. Since W and $W^\perp$ are subspaces, the sums $x_W + y_W$ and $x_{W^\perp} + y_{W^\perp}$ are in W and $W^\perp$, respectively. Therefore, the orthogonal decomposition of $x + y$ is $(x_W + y_W) + (x_{W^\perp} + y_{W^\perp})$, so
$$T(x+y) = (x+y)_W = x_W + y_W = T(x) + T(y).$$
Now let c be a scalar. Then cx_W is in W and $cx_{W^\perp}$ is in $W^\perp$, so the orthogonal decomposition of cx is $cx_W + cx_{W^\perp}$, and therefore,
$$T(cx) = (cx)_W = cx_W = cT(x).$$
Since T satisfies the two defining properties, Definition 3.3.1 in Section 3.3, it is a linear transformation.
2. See Example 7.3.2
3. See Example 7.3.3
4. For any x in $\mathbb{R}^n$ the vector $T(x)$ is in W, so $T \circ T(x) = T(T(x)) = T(x)$ by 2. Any vector x in W is in the range of T, because $T(x) = x$ for such vectors. On the other hand, for any vector x in $\mathbb{R}^n$ the output $T(x) = x_W$ is in W, so W is the range of T.

We compute the standard matrix of the orthogonal projection in the same way as for any other transformation, Theorem 3.3.1 in Section 3.3: by evaluating on the standard coordinate vectors. In this case, this means projecting the standard coordinate vectors onto the subspace.

Example 7.3.15: Matrix of a projection

Let L be the line in $\mathbb{R}^2$ spanned by the vector $u = \begin{pmatrix}3\\2\end{pmatrix}$, and define $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ by $T(x) = x_L$. Compute the standard matrix B for T.

Solution

The columns of B are $T(e_1) = (e_1)_L$ and $T(e_2) = (e_2)_L$. We have
$$\left.\begin{aligned}(e_1)_L &= \frac{u\cdot e_1}{u\cdot u}\quad u = \frac{3}{13}\begin{pmatrix}3\\2\end{pmatrix}\\ (e_2)_L &= \frac{u\cdot e_2}{u\cdot u}\quad u = \frac{2}{13}\begin{pmatrix}3\\2\end{pmatrix}\end{aligned}\right\} \implies B = \frac{1}{13}\begin{pmatrix}9 & 6\\ 6 & 4\end{pmatrix}.$$

Example 7.3.16: Matrix of a projection

Let L be the line in $\mathbb{R}^2$ spanned by the vector
$$u = \begin{pmatrix}-1\\1\\1\end{pmatrix},$$
and define $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ by $T(x) = x_L$. Compute the standard matrix B for T.

Solution

The columns of B are $T(e_1) = (e_1)_L$, $T(e_2) = (e_2)_L$, and $T(e_3) = (e_3)_L$. We have

$$\left.\begin{aligned} (e_1)_L &= \frac{u \cdot e_1}{u \cdot u} \quad u = \frac{-1}{3}\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \\ (e_2)_L &= \frac{u \cdot e_2}{u \cdot u} \quad u = \frac{1}{3}\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \\ (e_3)_L &= \frac{u \cdot e_3}{u \cdot u} \quad u = \frac{1}{3}\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \end{aligned}\right\} \implies B = \frac{1}{3}\begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix}.$$

Example 7.3.17: Matrix of a projection

Continuing with Example 7.3.11, let

$$W = \text{Span}\left\{\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right\},$$

and define $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ by $T(x) = x_W$. Compute the standard matrix B for T.

Solution

The columns of B are $T(e_1) = (e_1)_W$, $T(e_2) = (e_2)_W$, and $T(e_3) = (e_3)_W$. Let

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

To compute each $(e_i)_W$, we solve the matrix equation $A^TAc = A^Te_i$ for c, then use the equality $(e_i)_W = Ac$. First we note that

$$A^TA = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}; \qquad A^Te_i = \text{the } i\text{th column of } A^T = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix}.$$

For e_1, we form an augmented matrix and row reduce:

$$\left(\begin{array}{cc|c} 2 & 1 & 1 \\ 1 & 2 & 1 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & 1/3 \\ 0 & 1 & 1/3 \end{array}\right) \implies (e_1)_W = A\begin{pmatrix} 1/3 \\ 1/3 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}.$$

We do the same for e_2:

$$\left(\begin{array}{cc|c} 2 & 1 & 0 \\ 1 & 2 & 1 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & -1/3 \\ 0 & 1 & 2/3 \end{array}\right) \implies (e_1)_W = A\begin{pmatrix} -1/3 \\ 2/3 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

and for e_3:

$$\left(\begin{array}{cc|c} 2 & 1 & -1 \\ 1 & 2 & 0 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & -2/3 \\ 0 & 1 & 1/3 \end{array}\right) \implies (e_1)_W = A\begin{pmatrix} -2/3 \\ 1/3 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}.$$

It follows that

$$B = \frac{1}{3}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}.$$

In the previous Example 7.3.17, we could have used the fact that

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\}$$

forms a *basis* for W, so that

$$T(x) = x_W = \left[A(A^TA)^{-1}A^T\right]x \quad \text{for} \quad A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 0 \end{pmatrix}$$

by the Corollary 7.3.1. In this case, we have already expressed T as a matrix transformation with matrix $A(A^TA)^{-1}A^T$. See this Example 7.3.14

Note 7.3.2

Let W be a subspace of $\mathbb{R}^n$ with basis $v_1, v_2, \ldots, v_m$, and let A be the matrix with columns $v_1, v_2, \ldots, v_m$. Then the standard matrix for $T(x) = x_W$ is

$$A(A^TA)^{-1}A^T.$$

We can translate the above properties of orthogonal projections, Proposition 7.3.1, into properties of the associated standard matrix.

Proposition 7.3.2: Properties of Projection Matrices

Let W be a subspace of $\mathbb{R}^n$, define $T\colon \mathbb{R}^n \to \mathbb{R}^n$ by $T(x) = x_W$, and let B be the standard matrix for T. Then:

1. $\mathrm{Col}(B) = W$.
2. $\mathrm{Nul}(B) = W^\perp$.
3. $B^2 = B$.
4. If $W \neq \{0\}$, then 1 is an eigenvalue of B and the 1-eigenspace for B is W.
5. If $W \neq \mathbb{R}^n$, then 0 is an eigenvalue of B and the 0-eigenspace for B is $W^\perp$.
6. B is similar to the diagonal matrix with m ones and $n-m$ zeros on the diagonal, where $m = \dim(W)$.

Proof

The first four assertions are translations of properties 5, 3, 4, and 2 from Proposition 7.3.2 respectively, using this Note 3.1.1 in Section 3.1 and Theorem 3.4.1 in Section 3.4. The fifth assertion is equivalent to the second, by Fact 5.1.2 in Section 5.1.

For the final assertion, we showed in the proof of this Theorem 7.3.1 that there is a basis of $\mathbb{R}^n$ of the form $\{v_1, \ldots, v_m, v_{m+1}, \ldots, v_n\}$, where $\{v_1, \ldots, v_m\}$ is a basis for W and $\{v_{m+1}, \ldots, v_n\}$ is a basis for $W^\perp$. Each v_i is an eigenvector of B: indeed, for $i \leq m$ we have

$$Bv_i = T(v_i) = v_i = 1 \cdot v_i$$

because v_i is in W, and for $i > m$ we have

$$Bv_i = T(v_i) = 0 = 0 \cdot v_i$$

because v_i is in $W^\perp$. Therefore, we have found a basis of eigenvectors, with associated eigenvalues $1, \ldots, 1, 0, \ldots, 0$ (m ones and $n-m$ zeros). Now we use Theorem 5.4.1 in Section 5.4.

We emphasize that the properties of projection matrices, Proposition 7.3.2, would be very hard to prove in terms of matrices. By translating all of the statements into statements about linear transformations, they become much more transparent. For example, consider the projection matrix we found in Example 7.3.17. Just by looking at the matrix it is not at all obvious that when you square the matrix you get the same matrix back.

Example 7.3.18

Continuing with above Example 7.3.17, we showed that

$$B = \frac{1}{3}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}$$

is the standard matrix of the orthogonal projection onto

$$W = \text{Span}\left\{\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right\}.$$

One can verify by hand that $B^2 = B$ (try it!). We compute $W^\perp$ as the null space of

$$\begin{pmatrix} 1 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}.$$

The free variable is x_3, and the parametric form is $x_1 = x_3,\ x_2 = -x_3,$ so that

$$W^\perp = \text{Span}\left\{\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\right\}.$$

It follows that B has eigenvectors

$$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \qquad \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

with eigenvalues $1, 1, 0,$ respectively, so that

$$B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix}^{-1}.$$

Remark

As we saw in Example 7.3.18, if you are willing to compute bases for W and $W^\perp$, then this provides a third way of finding the standard matrix B for projection onto W: indeed, if $\{v_1, v_2, \ldots, v_m\}$ is a basis for W and $\{v_{m+1}, v_{m+2}, \ldots, v_n\}$ is a basis for $W^\perp$, then

$$B = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}\begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}\begin{pmatrix} | & | & & | \\ v_1 & v_1 & \cdots & v_n \\ | & | & & | \end{pmatrix}^{-1},$$

where the middle matrix in the product is the diagonal matrix with m ones and $n-m$ zeros on the diagonal. However, since you already have a basis for W, it is faster to multiply out the expression $A(A^TA)^{-1}A^T$ as in Corollary 7.3.1.

Remark: Reflections

Let W be a subspace of $\mathbb{R}^n$, and let x be a vector in $\mathbb{R}^n$. The *reflection* of x over W is defined to be the vector

$$\text{ref}_W(x) = x - 2x_{W^\perp}.$$

In other words, to find $\text{ref}_W(x)$ one starts at x, then moves to $x - x_{W^\perp} = x_W,$ then continues in the same direction one more time, to end on the opposite side of W.

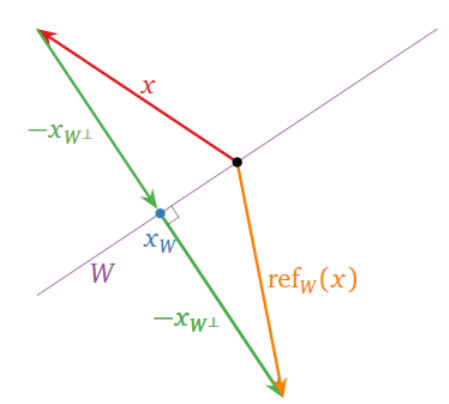


Figure 7.3.14

Since $x_{W^\perp} = x - x_W,$ we also have

$$\text{ref}_W(x) = x - 2(x - x_W) = 2x_W - x.$$

We leave it to the reader to check using the definition that:

1. $\text{ref}_W \circ \text{ref}_W = \text{Id}_{\mathbb{R}^n}$.
2. The 1-eigenspace of ref_W is W, and the -1-eigenspace of ref_W is $W^\perp$.
3. ref_W is similar to the diagonal matrix with $m = \dim(W)$ ones on the diagonal and $n - m$ negative ones.

This page titled 7.3: Orthogonal Projection is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **6.3: Orthogonal Projection** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

7.4: Orthogonal Sets and Gram-Schmidt Process

Learning Objectives

1. Understand which is the best method to use to compute an orthogonal projection in a given situation.
2. *Recipes:* an orthonormal set from an orthogonal set, Projection Formula, B-coordinates when B is an orthogonal set, Gram–Schmidt process.
3. *Vocabulary words:* ***orthogonal set***, ***orthonormal set***.

In this section, we give a formula for orthogonal projection that is considerably simpler than the one in Section 6.3, in that it does not require row reduction or matrix inversion. However, this formula, called the Projection Formula, only works in the presence of an *orthogonal* basis. We will also present the Gram–Schmidt process for turning an arbitrary basis into an orthogonal one.

Orthogonal Sets and the Projection Formula

Computations involving projections tend to be much easier in the presence of an *orthogonal* set of vectors.

Definition 7.4.1: Orthogonal and Orthonormal

A set of *nonzero* vectors $\{u_1,\ u_2, \cdots, u_m\}$ is called ***orthogonal*** if $u_i \cdot u_j = 0$ whenever $i \neq j$. It is ***orthonormal*** if it is orthogonal, and in addition $u_i \cdot u_i = 1$ for all $i = 1,\ 2, \cdots, m$.

In other words, a set of vectors is orthogonal if different vectors in the set are perpendicular to each other. An orthonormal set is an orthogonal set of unit vectors, Definition 6.1.3 in Section 6.1.

Example 7.4.1

The standard coordinate vectors in $\mathbb{R}^n$ always form an orthonormal set. For instance, in $\mathbb{R}^3$ we check that

$$\begin{pmatrix}1\\0\\0\end{pmatrix}\cdot\begin{pmatrix}0\\1\\0\end{pmatrix}=0 \quad \begin{pmatrix}1\\0\\0\end{pmatrix}\cdot\begin{pmatrix}0\\0\\1\end{pmatrix}=0 \quad \begin{pmatrix}0\\1\\0\end{pmatrix}\cdot\begin{pmatrix}0\\0\\1\end{pmatrix}=0.$$

Since $e_i \cdot e_i = 1$ for all $i = 1,\ 2,\ 3$, this shows that $\{e_1,\ e_2,\ e_3\}$ is orthonormal.

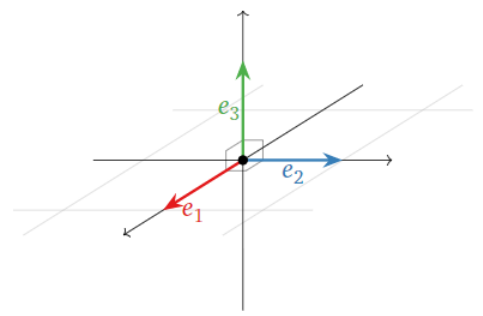


Figure 7.4.1

Example 7.4.2

Is this set orthogonal? Is it orthonormal?

$$\mathcal{B} = \left\{\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\-2\\1\end{pmatrix}, \begin{pmatrix}1\\0\\-1\end{pmatrix}\right\}$$

Solution

We check that

$$\begin{pmatrix}1\\1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\-2\\1\end{pmatrix}=0 \quad \begin{pmatrix}1\\1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\0\\-1\end{pmatrix}=0 \quad \begin{pmatrix}1\\-2\\1\end{pmatrix}\cdot\begin{pmatrix}1\\0\\-1\end{pmatrix}=0.$$

Therefore, $\mathcal{B}$ is orthogonal.

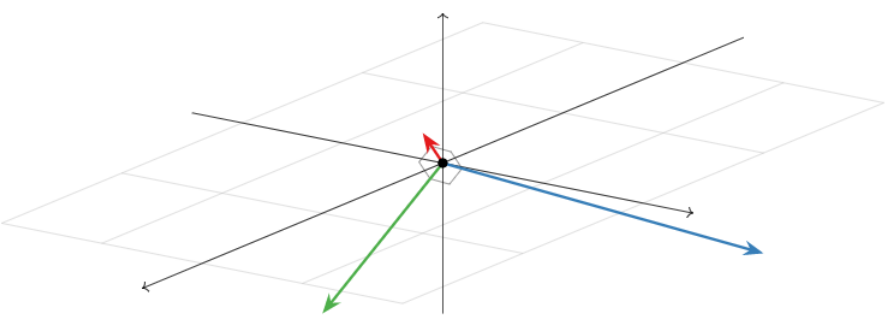


Figure 7.4.2

The set $\mathcal{B}$ is not orthonormal because, for instance,

$$\begin{pmatrix}1\\1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\1\\1\end{pmatrix}=3\neq 1.$$

However, we can make it orthonormal by replacing each vector by the unit vector in the direction of, Fact 6.1.3 in Section 6.1, each vector:

$$\left\{\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix},\ \frac{1}{\sqrt{6}}\begin{pmatrix}1\\-2\\1\end{pmatrix},\ \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix}\right\}$$

is orthonormal.

We saw in the previous example that it is easy to produce an orthonormal set of vectors from an orthogonal one by replacing each vector with the unit vector in the same direction.

Recipe: An Orthonormal Set from an Orthogonal Set

If $\{v_1,\ v_2,\cdots,v_m\}$ is an orthogonal set of vectors, then

$$\left\{\frac{v_1}{||v_1||},\ \frac{v_2}{||v_2||},\cdots,\frac{v_m}{||v_m||}\right\}$$

is an orthonormal set.

Example 7.4.3

Let $a,\ b$ be scalars, and let

$$u_1=\begin{pmatrix}a\\b\end{pmatrix} \quad u_2=\begin{pmatrix}-b\\a\end{pmatrix}.$$

Is $\mathcal{B}=\{u_1,\ u_2\}$ orthogonal?

Solution

We only have to check that

$$\begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} -b \\ a \end{pmatrix} = -ab + ab = 0.$$

Therefore, $\{u_1,\ u_2\}$ is orthogonal, *unless* $a = b = 0$.

Example 7.4.4: Non-Example

Is this set orthogonal?

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} \right\}$$

Solution

This set is not orthogonal because

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} = 1 - 1 - 1 = -1 \neq 0.$$

We will see how to product an orthogonal set from $\mathcal{B}$ in the subsection: The Gram-Schmidt Process.

A nice property enjoyed by orthogonal sets is that they are automatically linearly independent.

Fact 7.4.1

An orthogonal set is linearly independent. Therefore, it is a basis for its span.

Proof

Suppose that $\{u_1,\ u_2, \cdots, u_m\}$ is orthogonal. We need to show that the equation

$$c_1 u_1 + c_2 u_2 + \cdots + c_m u_m = 0$$

has only the trivial solution $c_1 = c_2 = \cdots = c_m = 0$. Taking the dot product of both sides of this equation with u1 gives

$$\begin{aligned} 0 = u_1 \cdot 0 \ &= u_1 \cdot (c_1 u_1 + c_2 u_2 + \cdots + c_m u_m) \\ &= c_1 (u_1 \cdot u_1) + c_2 (u_1 \cdot u_2) + \cdots + c_m (u_1 \cdot u_m) \\ &= c_1 (u_1 \cdot u_1) \end{aligned}$$

because $u_1 \cdot u_i = 0$ for $i > 1$. Since $u \neq 0$ we have $u_1 \cdot u_1 \neq 0$, so $c_1 = 0$. Similarly, taking the dot product with u_i shows that each $c_i = 0$, as desired.

One advantage of working with orthogonal sets is that it gives a simple formula for the orthogonal projection, Definition 6.3.2 in Section 6.3, of a vector.

Theorem 7.4.1: Projection Formula

Let W be a subspace of $\mathbb{R}^n$, and let $\{u_1,\ u_2, \cdots, u_m\}$ be an orthogonal basis for W. Then for any vector x in $\mathbb{R}^n$, the orthogonal projection of x onto W is given by the formula

$$x_W = \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2} u_2 + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m} u_m.$$

Proof

Let

$$y = \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2} u_2 + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m} u_m.$$

This vector is contained in W because it is a linear combination of $u_1,\ u_2, \cdots, u_m$. Hence we just need to show that $x - y$ is in $W^\perp$, i.e., that $u_i \cdot (x - y) = 0$ for each $i = 1,\ 2, \cdots, m$. For u_1, we have

$$\begin{aligned} u_1 \cdot (x-y) &= u_1 \cdot \left(x - \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 - \frac{x \cdot u_2}{u_2 \cdot u_2} u_2 - \cdots - \frac{x \cdot u_m}{u_m \cdot u_m} u_m \right) \\ &= u_1 \cdot x - \frac{x \cdot u_1}{u_1 \cdot u_1}(u_1 \cdot u_1) - 0 - \cdots - 0 \\ &= 0. \end{aligned}$$

A similar calculation shows that $u_i \cdot (x - y) = 0$ for each i, so $x - y$ is in $W^\perp$, as desired.

If $\{u_1,\ u_2, \cdots, u_m\}$ is an ortho*normal* basis for W, then the denominators $u_i \cdot u_i = 1$ go away, so the projection formula becomes even simpler:

$$x_W = (x \cdot u_1)u_1 + (x \cdot u_2)u_2 + \cdots + (x \cdot u_m)u_m.$$

Example 7.4.5

Suppose that $L = \text{Span}\{u\}$ is a line. The set $\{u\}$ is an orthogonal basis for L, so the Projection Formula says that for any vector x, we have

$$x_L = \frac{x \cdot u}{u \cdot u} u,$$

as in Example 6.3.7 in Section 6.3. See also Example 6.3.8 in Section 6.3 and Example 6.3.9 in Section 6.3.

Suppose that $\{u_1,\ u_2, \cdots, u_m\}$ is an orthogonal basis for a subspace W, and let $L_i = \text{Span}\{u_i\}$ for each $i = 1,\ 2, \cdots, m$. Then we see that for any vector x, we have

$$\begin{aligned} x_W &= \frac{x \cdot u_1}{u_1 \cdot u_1} u_2 + \frac{x \cdot u_2}{u_2 \cdot u_2} u_2 + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m} u_m \\ &= x_{L_1} + x_{L_2} + \cdots + x_{L_m} \end{aligned}$$

In other words, for an *orthogonal* basis, the projection of x onto W is the sum of the projections onto the lines spanned by the basis vectors. In this sense, projection onto a line is the most important example of an orthogonal projection.

Example 7.4.6: Projection onto the xy-Plane

Continuing with Example 6.3.1 in Section 6.3 and Example 6.3.10 in Section 6.3, use the projection formula, Theorem 7.4.1 to compute the orthogonal projection of a vector onto the xy-plane in $\mathbb{R}^3$.

Solution

A basis for the xy-plane is given by the two standard coordinate vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

The set $\{e_1,\ e_2\}$ is orthogonal, so for any vector $x = (x_1,\ x_2,\ x_3)$, we have

$$x_W = \frac{x \cdot e_1}{e_1 \cdot e_1} e_1 + \frac{x \cdot e_2}{e_2 \cdot e_2} e_2 = x_1 e_1 + x_2 e_2 = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}.$$

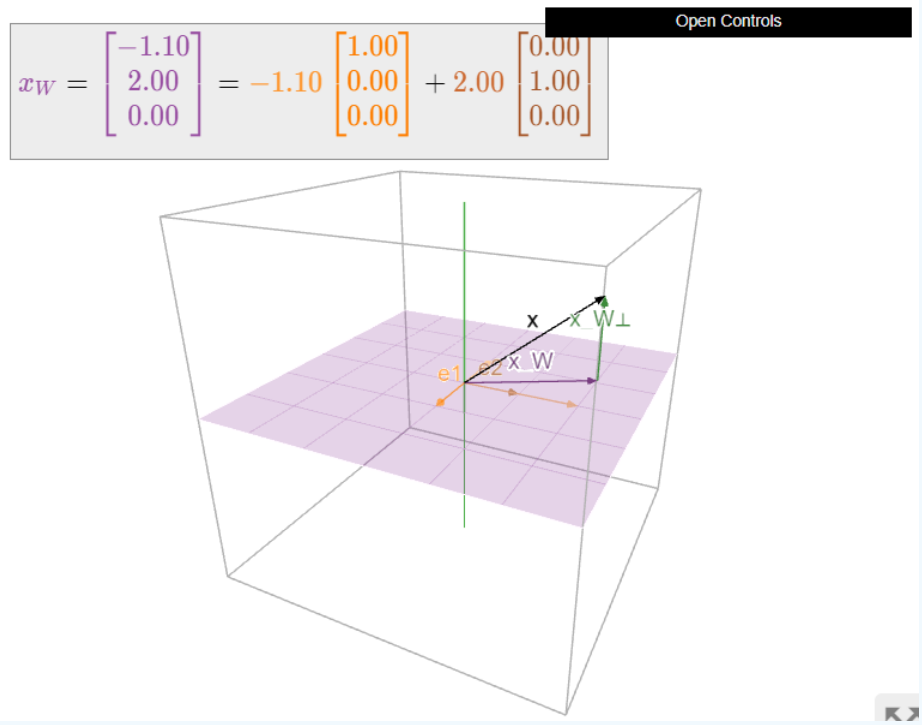


Figure 7.4.3: Orthogonal projection of a vector onto the xy-plane in $\mathbb{R}^3$. Note that x_W is the sum of the projections of x onto the e_1- and e_2-coordinate axes (shown in orange and brown, respectively).

Example 7.4.7: Projection onto a Plane in $\mathbb{R}^3$

Let

$$W = \text{Span}\left\{\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right\} \quad x = \begin{pmatrix} 2 \\ 3 \\ -2 \end{pmatrix}.$$

Find x_W and $x_{W^\perp}$.

Solution

The vectors

$$u_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \quad u_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

are orthogonal, so we can use the Projection Formula:

$$x_W = \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2} u_2 = \frac{4}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + \frac{3}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}.$$

Then we have

$$x_{W^\perp} = x - x_W = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}.$$

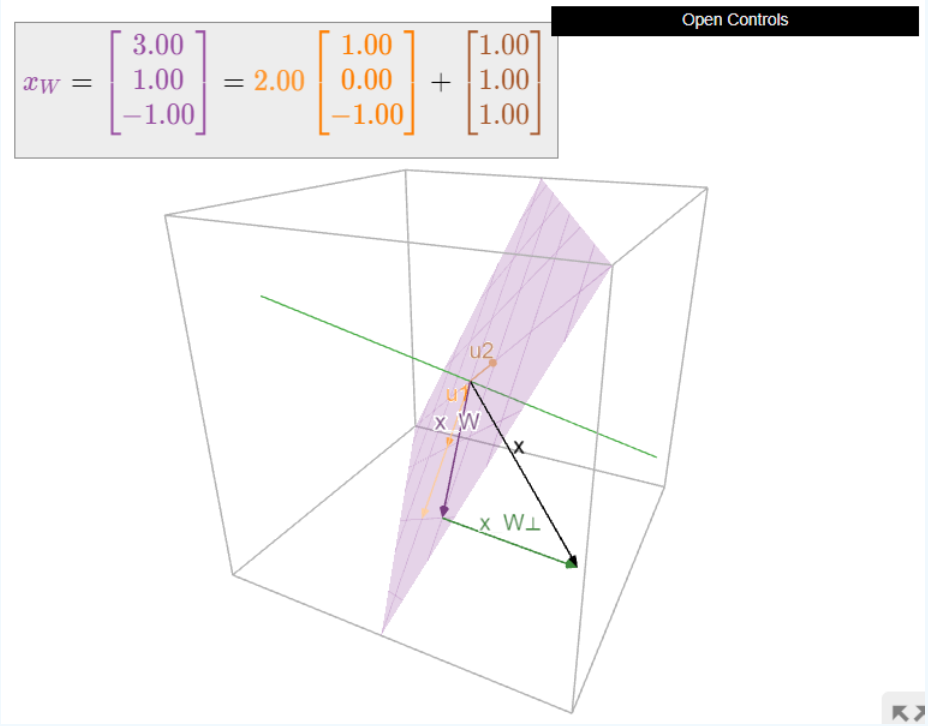


Figure 7.4.4: Orthogonal projection of a vector onto the plane W. Note that x_W is the sum of the projections of x onto the lines spanned by u_1 and u_2 (shown in orange and brown, respectively).

Example 7.4.8: Projection onto a 3-Space in $\mathbb{R}^4$

Let

$$W = \text{Span}\left\{\begin{pmatrix}1\\0\\-1\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\\-1\end{pmatrix}, \begin{pmatrix}1\\1\\1\\1\end{pmatrix}\right\} \quad x = \begin{pmatrix}0\\1\\3\\4\end{pmatrix}.$$

Compute x_W, and find the distance from x to W.

Solution

The vectors

$$u_1 = \begin{pmatrix}1\\0\\-1\\0\end{pmatrix} \quad u_2 = \begin{pmatrix}0\\1\\0\\-1\end{pmatrix} \quad u_3 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}$$

are orthogonal, so we can use the Projection Formula:

$$\begin{aligned} x_W &= \frac{x\cdot u_1}{u_1\cdot u_1}u_1+\frac{x\cdot u_2}{u_2\cdot u_2}u_2+\frac{x\dot{u}_3}{u_3\cdot u_3}u_3 \\ &= \frac{-3}{2}\begin{pmatrix}1\\0\\-1\\0\end{pmatrix}+\frac{-3}{2}\begin{pmatrix}0\\1\\0\\-1\end{pmatrix}+\frac{8}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix}=\frac{1}{2}\begin{pmatrix}1\\1\\7\\7\end{pmatrix} \\ x_{W^\perp} &= x-x_W=\frac{1}{2}\begin{pmatrix}-1\\1\\-1\\1\end{pmatrix}. \end{aligned}$$

The distance from x to W is

$$||x_{W^\perp}|| = \frac{1}{2}\sqrt{1+1+1+1} = 1.$$

Now let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\mathcal{B}=\{v_1,\ v_2,\cdots,v_m\}$, and let x be a vector in W. Then $x=x_W$, so by the projection formula, Theorem 7.4.1, we have

$$x = x_W = \frac{x\cdot u_1}{u_1\cdot u_1}u_1+\frac{x\cdot u_2}{u_2\cdot u_2}u_2+\cdots+\frac{x\cdot u_m}{u_m\cdot u_m}u_m.$$

This gives us a way of expressing x as a linear combination of the basis vectors in $\mathcal{B}$: we have computed the $\mathcal{B}$-coordinates, Definition 2.8.1 in Section 2.8, of x without row reducing!

Recipe: $\mathcal{B}$-Coordinates when $\mathcal{B}$ is an Orthogonal Set

Let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\mathcal{B}=\{v_1,\ v_2,\cdots,v_m\}$ and let x be a vector in W. Then

$$[x]_{\mathcal{B}} = \left(\frac{x\cdot u_1}{u_1\cdot u_1},\ \frac{x\cdot u_2}{u_2\cdot u_2},\cdots,\frac{x\cdot u_m}{u_m\cdot u_m}\right).$$

As with orthogonal projections, if $\{u_1,\ u_2,\cdots,u_m\}$ is an ortho*normal* basis of W, then the formula is even simpler:

$$[x]_{\mathcal{B}} = (x\cdot u_1,\ x\cdot u_2,\cdots,x\cdot u_m).$$

Example 7.4.9: Computing Coordinates with Respect to an Orthogonal Basis

Find the $\mathcal{B}$-coordinates of x, where

$$\mathcal{B}=\left\{\begin{pmatrix}1\\2\end{pmatrix},\begin{pmatrix}-4\\2\end{pmatrix}\right\}\quad x=\begin{pmatrix}0\\3\end{pmatrix}.$$

Solution

Since

$$u_1=\begin{pmatrix}1\\2\end{pmatrix}\quad u_2=\begin{pmatrix}-4\\2\end{pmatrix}$$

form an *orthogonal* basis of $\mathbb{R}^2$, we have

$$[x]_{\mathcal{B}} = \left(\frac{x\cdot u_1}{u_1\cdot u_1},\frac{x\cdot u_2}{u_2\cdot u_2}\right) = \left(\frac{3\cdot 2}{1^2+2^2},\frac{3\cdot 2}{(-4)^2+2^2}\right) = \left(\frac{6}{5},\frac{3}{10}\right).$$

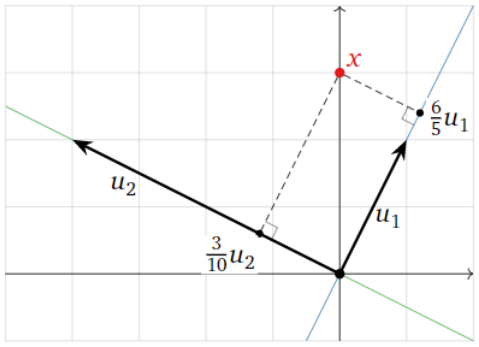


Figure 7.4.5

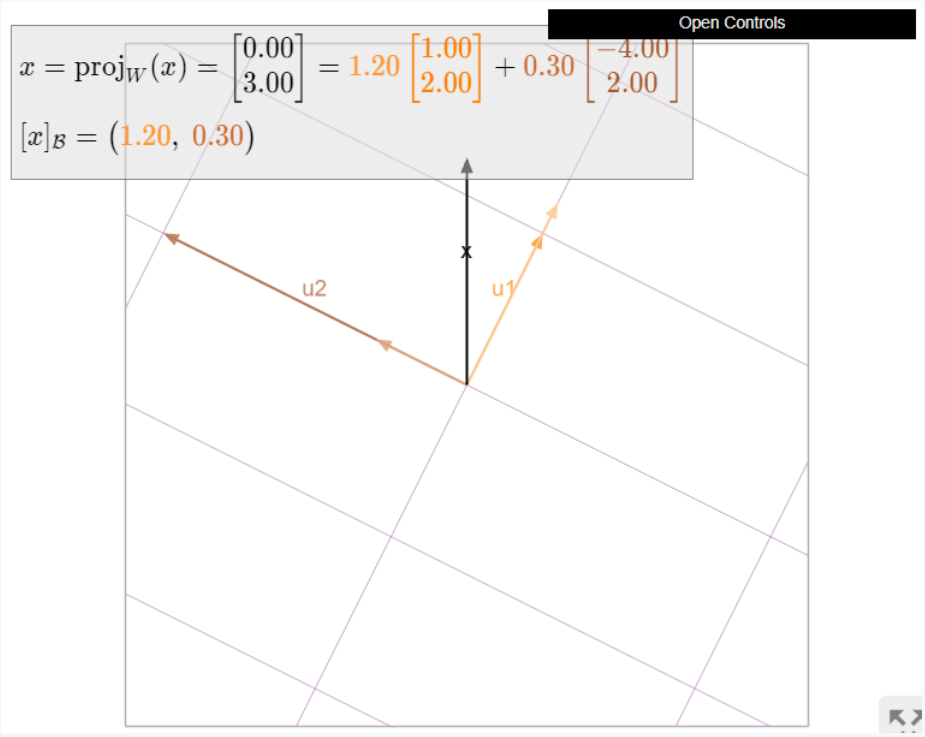


Figure 7.4.6: Computing $\mathcal{B}$-coordinates using the Projection Formula.

The following example shows that the Projection Formula does in fact require an orthogonal basis.

Example 7.4.10: Non-Example: A Non-Orthogonal Basis

Consider the basis $\mathcal{B} = \{v_1,\ v_2\}$ of $\mathbb{R}^2$, where

$$v_1 = \begin{pmatrix} 2 \\ -1/2 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

This is not orthogonal because $v_1 \cdot v_2 = 2 - 1 = 1 \neq 0$. Let $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Let us try to compute $x = x_{\mathbb{R}^2}$ using the Projection Formula with respect to the basis $\mathcal{B}$:

$$x_{\mathbb{R}^2} = \frac{x \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{x \cdot v_2}{v_2 \cdot v_2} v_2 = \frac{3/2}{17/4}\begin{pmatrix} 2 \\ -1/2 \end{pmatrix} + \frac{3}{5}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 111/85 \\ 87/85 \end{pmatrix} \neq x.$$

Since $x = x_{\mathbb{R}^2}$, we see that the Projection Formula does not compute the orthogonal projection in this case. Geometrically, the projections of x onto the lines spanned by v_1 and v_2 do not sum to x, as we can see from the picture.

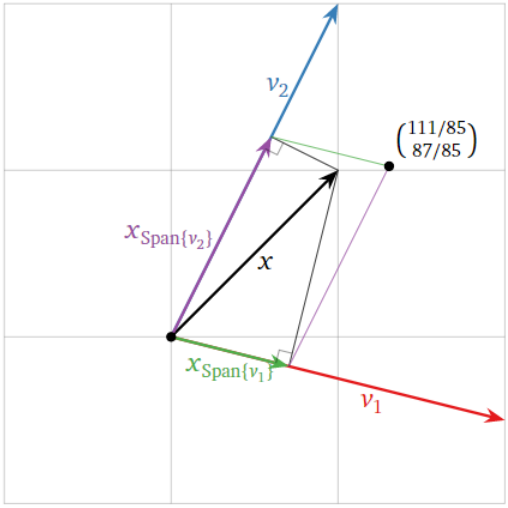


Figure 7.4.7

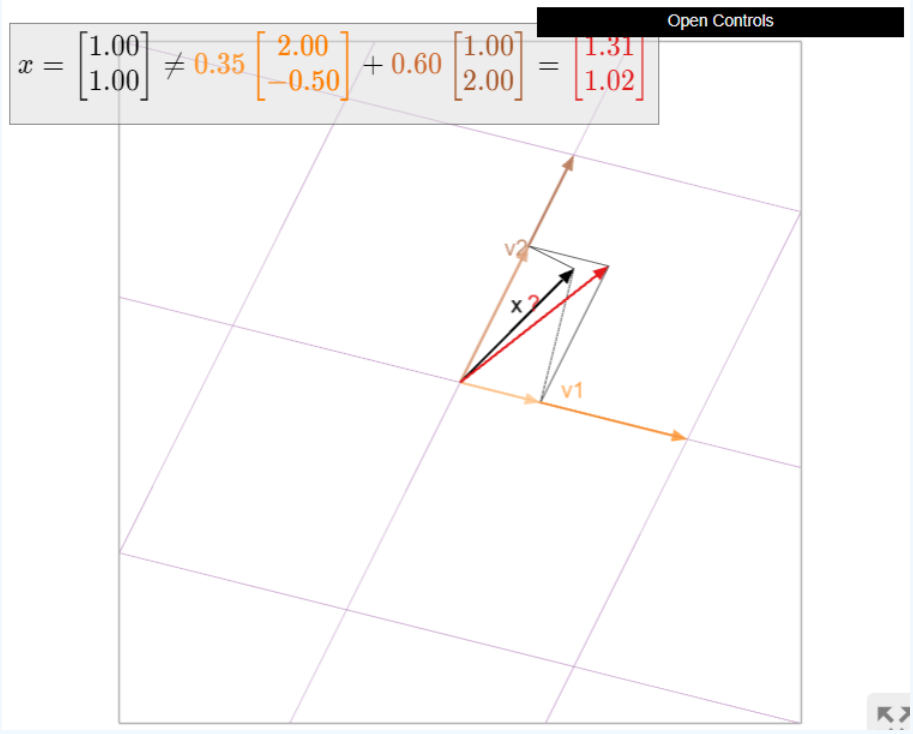


Figure 7.4.8: When v_1 and v_2 are not orthogonal, then $x_{\mathbb{R}^2} = x$ is not necessarily equal to the sum (red) of the projections (orange and brown) of x onto the lines spanned by v_1 and v_2.

Note 7.4.1

You need an orthogonal basis to use the Projection Formula.

The Gram-Schmidt Process

We saw in the previous subsection that orthogonal projections and $\mathcal{B}$-coordinates are much easier to compute in the presence of an *orthogonal* basis for a subspace. In this subsection, we give a method, called the *Gram–Schmidt Process*, for computing an orthogonal basis of a subspace.

Theorem 7.4.2: The Gram-Schmidt Process

Let $v_1,\ v_2, \cdots, v_m$ be a basis for a subspace W of $\mathbb{R}^n$. Define:

1. $u_1 = v_1$
2. $u_2 = (v_2)_{\text{Span}\{u_1\}^\perp} = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1$
3. $u_3 = (v_3)_{\text{Span}\{u_1, u_2\}^\perp} = v_3 - \frac{v_3 \cdot u_1}{u_1 \cdot u_1} u_1 - \frac{v_3 \cdot u_2}{u_2 \cdot u_2} u_2$

$\vdots$

M. $u_m = (v_m)_{\text{Span}\{u_1, u_2, \cdots, u_{m-1}\}^\perp} = v_m - \sum_{i=1}^{m-1} \frac{v_m \cdot u_i}{u_i \cdot u_i} u_i$.

Then $\{u_1,\ u_2, \cdots, u_m\}$ is an orthogonal basis for the same subspace W.

Proof

First we claim that each u_i is in W, and in fact that u_i is in $\text{Span}\{v_1,\ v_2, \cdots, v_i\}$. Clearly $u_1 = v_1$ is in $\text{Span}\{v_1\}$. Then u_2 is a linear combination of u_1 and v_2, which are both in $\text{Span}\{v_1,\ v_2\}$, so u_2 is in $\text{Span}\{v_1,\ v_2\}$ as well. Similarly, u_3 is a linear combination of $u_1,\ u_2$, and v_3, which are all in $\text{Span}\{v_1,\ v_2,\ v_3\}$, so u_3 is in $\text{Span}\{v_1,\ v_2,\ v_3\}$. Continuing in this way, we see that each u_i is in $\text{Span}\{v_1,\ v_2, \cdots, v_i\}$.

Now we claim that $\{u_1,\ u_2, \cdots, u_m\}$ is an orthogonal set. Let $1 \le i < j \le m$. Then $u_j = (v_j)_{\text{Span}\{u_1,\ u_2, \cdots, u_{j-1}\}^\perp}$, so by definition u_j is orthogonal to every vector in $\text{Span}\{u_1,\ u_2, \cdots, u_{j-1}\}$. In particular, u_j is orthogonal to u_i.

We still have to prove that each u_i is nonzero. Clearly $u_1 = v_1 \neq 0$. Suppose that $u_i = 0$. Then $(v_i)_{\text{Span}\{u_1,\ u_2, \cdots, u_{i-1}\}^\perp} = 0$, which means that v_i is in $\text{Span}\{u_1,\ u_2, \cdots, u_{i-1}\}$. But each $u_1,\ u_2, \cdots, u_{i-1}$ is in $\text{Span}\{v_1,\ v_2, \cdots, v_{i-1}\}$ by the first paragraph, so v_i is in $\text{Span}\{v_1,\ v_2, \cdots, v_{i-1}\}$. This contradicts the increasing span criterion Theorem 2.5.2 in Section 2.5; therefore, u_i must be nonzero.

The previous two paragraphs justify the use of the projection formula, Theorem 7.4.1, in the equalities

$$(v_i)_{\text{Span}\{u_1,\ u_2, \cdots, u_{i-1}\}^\perp} = v_i - (v_i)_{\text{Span}\{u_1,\ u_2, \cdots, u_{i-1}\}} = v_i - \sum_{j=1}^{i-1} \frac{v_i \cdot u_j}{u_j \cdot u_j} u_j$$

in the statement of the theorem.

Since $\{u_1,\ u_2, \cdots, u_m\}$ is an orthogonal set, it is linearly independent. Thus it is a set of m linearly independent vectors in W, so it is a basis for W by the basis theorem, Theorem 2.7.3 in Section 2.7. Similarly, for every i, we saw that the set $\{u_1,\ u_2, \cdots, u_i\}$ is contained in the i-dimensional subspace $\text{Span}\{v_1,\ v_2, \cdots, v_i\}$, so $\{u_1,\ u_2, \cdots, u_i\}$ is an orthogonal basis for $\text{Span}\{v_1,\ v_2, \cdots, v_i\}$.

Example 7.4.11: Two Vectors

Find an orthogonal basis $\{u_1,\ u_2\}$ for $W = \text{Span}\{v_1,\ v_2\}$, where

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Solution

We run Gram–Schmidt: first take $u_1 = v_1$, then

$$u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Then $\{u_1,\ u_2\}$ is an orthogonal basis for W: indeed, it is clear that $u_1 \cdot u_2 = 0$.

Geometrically, we are simply replacing v_2 with the part of v_2 that is perpendicular to line $L_1 = \text{Span}\{v_1\}$:

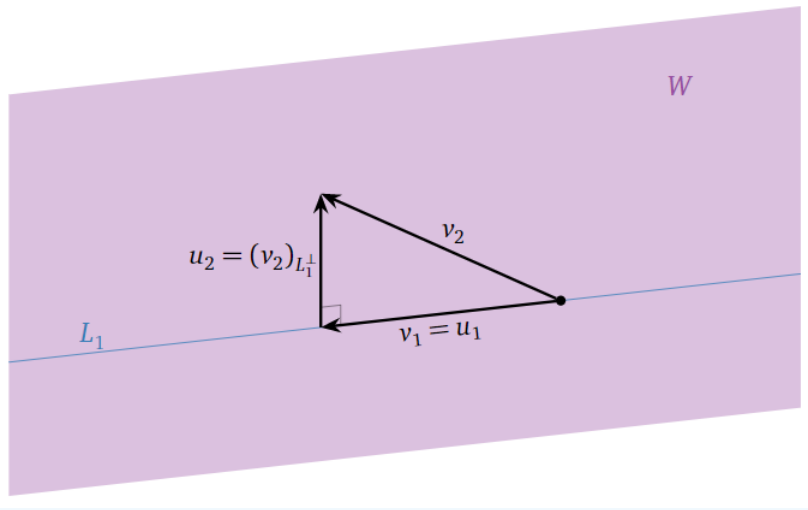


Figure 7.4.9

Example 7.4.12: Three Vectors

Find an orthogonal basis $\{u_1,\ u_2,\ u_3\}$ for $W = \text{Span}\{v_1,\ v_2,\ v_3\} = \mathbb{R}^3$, where

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad v_3 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}.$$

Solution

We run Gram-Schmidt:

1. $u_1 = v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$
2. $u_2 = v_2 = \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$
3. $u_3 = v_3 - \dfrac{v_3 \cdot u_1}{u_1 \cdot u_1} u_1 - \dfrac{v_3 \cdot u_2}{u_2 \cdot u_2} u_2$
$= \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} - \dfrac{4}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \dfrac{1}{1} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}.$

Then $\{u_1,\ u_2,\ u_3\}$ is an orthogonal basis for W: indeed, we have

$$u_1 \cdot u_2 = 0 \quad u_1 \cdot u_3 = 0 \quad u_2 \cdot u_3 = 0.$$

Geometrically, once we have u_1 and u_2, we replace v_3 by the part that is orthogonal to $W_2 = \text{Span}\{u_1,\ u_2\} = \text{Span}\{v_1,\ v_2\}$:

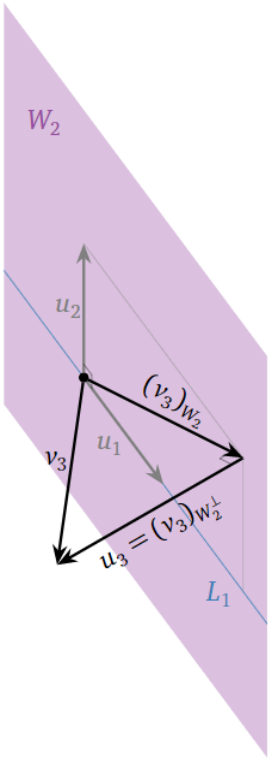


Figure 7.4.10

Example 7.4.13: Three Vectors in $\mathbb{R}^4$

Find an orthogonal basis $\{u_1,\ u_2,\ u_3\}$ for $W = \text{Span}\{v_1,\ v_2,\ v_3\}$, where

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad v_2 = \begin{pmatrix} -1 \\ 4 \\ 4 \\ -1 \end{pmatrix} \quad v_3 = \begin{pmatrix} 4 \\ -2 \\ -2 \\ 0 \end{pmatrix}.$$

Solution

We run Gram–Schmidt:

1. $u_1 = v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$

2. $u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1 = \begin{pmatrix} -1 \\ 4 \\ 4 \\ -1 \end{pmatrix} - \frac{6}{4} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -5/2 \\ 5/2 \\ 5/2 \\ -5/2 \end{pmatrix}$

3. $u_3 = v_3 - \dfrac{v_3 \cdot u_1}{u_1 \cdot u_1} u_1 - \dfrac{v_3 \cdot u_2}{u_2 \cdot u_2} u_2$

$$= \begin{pmatrix} 4 \\ -2 \\ -2 \\ 0 \end{pmatrix} - \frac{0}{24} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{-20}{25} \begin{pmatrix} -5/2 \\ 5/2 \\ 5/2 \\ -5/2 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \\ -2 \end{pmatrix}.$$

Then $\{u_1,\ u_2,\ u_3\}$ is an orthogonal basis for W.

We saw in the proof of the Gram–Schmidt Process that for every i between 1 and m, the set $\{u_1,\ u_2, \cdots, u_i\}$ is a an orthogonal basis for $\text{Span}\{v_1,\ v_2, \cdots, v_i\}$.

If we had started with a spanning set $\{v_1,\ v_2, \cdots, v_m\}$ which is linearly dependent, then for some i, the vector v_i is in $\text{Span}\{v_1,\ v_2, \cdots, v_{i-1}\}$ by the increasing span criterion, Theorem 2.5.2 in Section 2.5. Hence

$$0 = (v_i)_{\text{Span}\{v_1,\ v_2, \cdots, v_{i-1}\}^\perp} = (v_i)_{\text{Span}\{u_1,\ u_2, \cdots u_{i-1}\}^\perp} = u_i.$$

Note 7.4.2

You can use the Gram–Schmidt Process to produce an orthogonal basis from any spanning set: if some $u_i = 0$, just throw away u_i and v_i, and continue.

This page titled 7.4: Orthogonal Sets and Gram-Schmidt Process is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Dan Margalit & Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- **6.4: Orthogonal Sets** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

7.5: The Method of Least Squares

Learning Objectives

1. Learn examples of best-fit problems.
2. Learn to turn a best-fit problem into a least-squares problem.
3. *Recipe:* find a least-squares solution (two ways). solution.
4. *Picture:* geometry of a least-squares solution.
5. *Vocabulary words:* ***least-squares solution***.

In this section, we answer the following important question:

> *Suppose that $Ax = b$ does not have a solution. What is the best approximate solution?*

For our purposes, the best approximate solution is called the *least-squares solution*. We will present two methods for finding least-squares solutions, and we will give several applications to best-fit problems.

Least-Squares Solutions

We begin by clarifying exactly what we will mean by a "best approximate solution" to an inconsistent matrix equation $Ax = b$.

Definition 7.5.1: Least-Squares Solution

Let A be an $m \times n$ matrix and let b be a vector in $\mathbb{R}^m$. A ***least-squares solution*** of the matrix equation $Ax = b$ is a vector $\hat{x}$ in $\mathbb{R}^n$ such that

$$\text{dist}(b,\ A\hat{x}) \leq \text{dist}(b,\ Ax)$$

for all other vectors x in $\mathbb{R}^n$.

Recall that $\text{dist}(v, w) = \|v - w\|$ is the distance, Definition 6.1.2 in Section 6.1, between the vectors v and w. The term "least squares" comes from the fact that $\text{dist}(b, Ax) = \|b - A\hat{x}\|$ is the square root of the sum of the squares of the entries of the vector $b - A\hat{x}$. So a least-squares solution minimizes the sum of the squares of the differences between the entries of $A\hat{x}$ and b. In other words, a least-squares solution solves the equation $Ax = b$ as closely as possible, in the sense that the sum of the squares of the difference $b - Ax$ is minimized.

Least Squares: Picture

Suppose that the equation $Ax = b$ is inconsistent. Recall from Note 2.3.6 in Section 2.3 that the column space of A is the set of all other vectors c such that $Ax = c$ is consistent. In other words, $\text{Col}(A)$ is the set of all vectors of the form Ax. Hence, the closest vector, Note 6.3.1 in Section 6.3, of the form Ax to b is the orthogonal projection of b onto $\text{Col}(A)$. This is denoted $b_{\text{Col}(A)}$, following Definition 6.3.1 in Section 6.3.

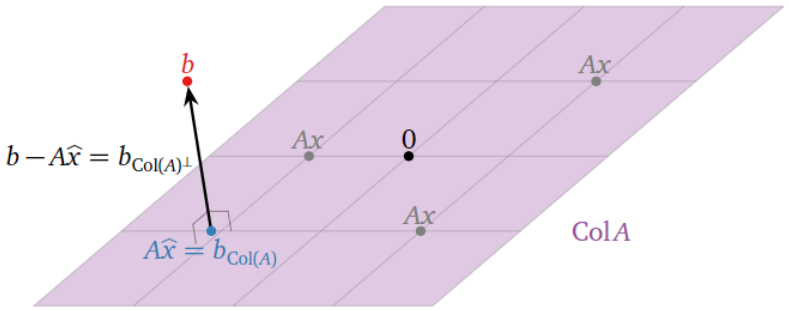


Figure 7.5.1

Note 7.5.1

A least-squares solution of $Ax = b$ is a solution $\hat{x}$ of the consistent equation $Ax = b_{\text{Col}(A)}$

Note 7.5.2

If $Ax = b$ is consistent, then $b_{\text{Col}(A)} = b$, so that a least-squares solution is the same as a usual solution.

Where is $\hat{x}$ in this picture? If $v_1, v_2, \ldots, v_n$ are the columns of A, then

$$A\hat{x} = A\begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_n \end{pmatrix} = \hat{x}_1 v_1 + \hat{x}_2 v_2 + \cdots + \hat{x}_n v_n.$$

Hence the entries of $\hat{x}$ are the "coordinates" of $b_{\text{Col}(A)}$ with respect to the spanning set $\{v_1, v_2, \ldots, v_m\}$ of $\text{Col}(A)$. (They are honest $\mathcal{B}$-coordinates if the columns of A are linearly independent.)

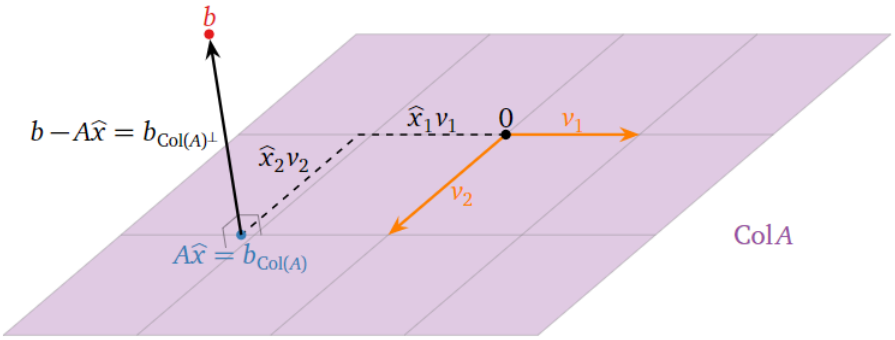


Figure 7.5.2

$$-1.10\, v_1 + 1.00\, v_2 = b_{\text{Col}(A)}$$

$$A\hat{x} = \begin{bmatrix} 0.00 & 1.10 \\ 1.00 & 0.00 \\ 0.00 & -0.20 \end{bmatrix} \begin{bmatrix} -1.10 \\ 1.00 \end{bmatrix} = \begin{bmatrix} 1.10 \\ -1.10 \\ -0.20 \end{bmatrix} = b_{\text{Col}(A)}$$

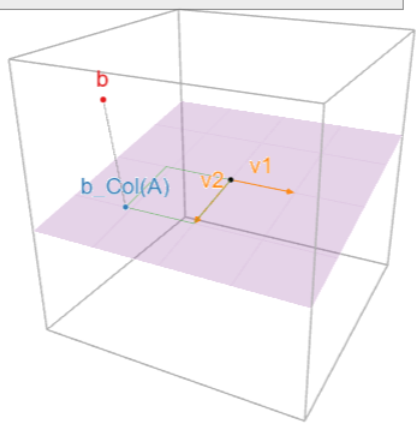


Figure 7.5.3: The violet plane is $\text{Col}(A)$. The closest that Ax can get to b is the closest vector on $\text{Col}(A)$ to b, which is the orthogonal projection $b_{\text{Col}(A)}$ (in blue). The vectors v_1, v_2 are the columns of A, and the coefficients of $\hat{x}$ are the lengths of the green lines. Click and drag b to move it.

We learned to solve this kind of orthogonal projection problem in Section 6.3.

Theorem 7.5.1

Let A be an $m \times n$ matrix and let b be a vector in $\mathbb{R}^m$. The least-squares solutions of $Ax = b$ are the solutions of the matrix equation

$$A^T Ax = A^T b$$

Proof

By Theorem 6.3.2 in Section 6.3, if $\hat{x}$ is a solution of the matrix equation $A^T Ax = A^T b$, then $A\hat{x}$ is equal to $b_{\mathrm{Col}(A)}$. We argued above that a least-squares solution of $Ax = b$ is a solution of $Ax = b_{\mathrm{Col}(A)}$.

In particular, finding a least-squares solution means solving a consistent system of linear equations. We can translate the above theorem into a recipe:

Recipe 1: Compute a Least-Squares Solution

Let A be an $m \times n$ matrix and let b be a vector in $\mathbb{R}^n$. Here is a method for computing a least-squares solution of $Ax = b$:

1. Compute the matrix $A^T A$ and the vector $A^T b$.
2. Form the augmented matrix for the matrix equation $A^T Ax = A^T b$, and row reduce.
3. This equation is always consistent, and any solution $\hat{x}$ is a least-squares solution.

To reiterate: once you have found a least-squares solution $\hat{x}$ of $Ax = b$, then $b_{\mathrm{Col}(A)}$ is equal to $A\hat{x}$.

Example 7.5.1

Find the least-squares solutions of $Ax = b$ where:

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

What quantity is being minimized?

Solution

We have

$$A^T A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix}$$

and

$$A^T b = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \end{pmatrix}.$$

We form an augmented matrix and row reduce:

$$\left(\begin{array}{cc|c} 5 & 3 & 0 \\ 3 & 3 & 6 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & -3 \\ 0 & 1 & 5 \end{array}\right).$$

Therefore, the only least-squares solution is $\hat{x} = \begin{pmatrix} -3 \\ 5 \end{pmatrix}$.

This solution minimizes the distance from $A\hat{x}$ to b, i.e., the sum of the squares of the entries of $b - A\hat{x} = b - b_{\mathrm{Col}(A)} = b_{\mathrm{Col}(A)^\perp}$. In this case, we have

$$||b - A\hat{x}|| = \left|\left|\begin{pmatrix}6\\0\\0\end{pmatrix} - \begin{pmatrix}5\\2\\-1\end{pmatrix}\right|\right| = \left|\left|\begin{pmatrix}1\\-2\\1\end{pmatrix}\right|\right| = \sqrt{1^2+(-2)^2+1^2} = \sqrt{6}.$$

Therefore, $b_{\text{Col}(A)} = A\hat{x}$ is $\sqrt{6}$ units from b.

In the following picture, v_1, v_2 are the columns of A:

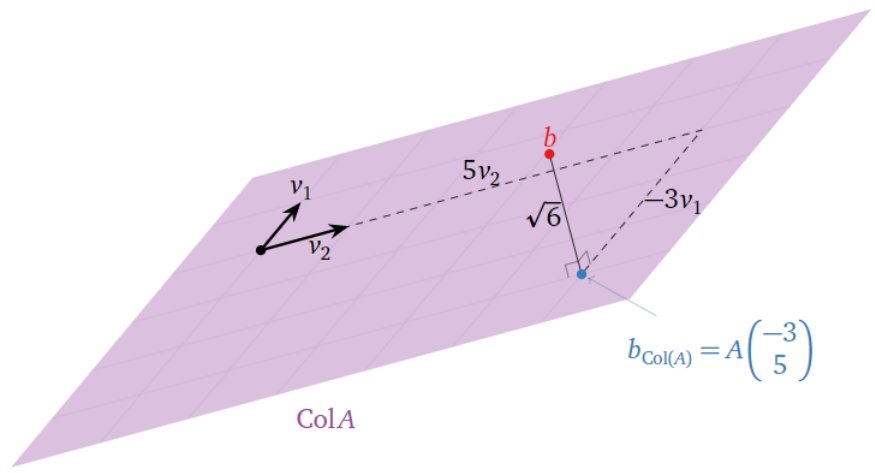


Figure 7.5.4

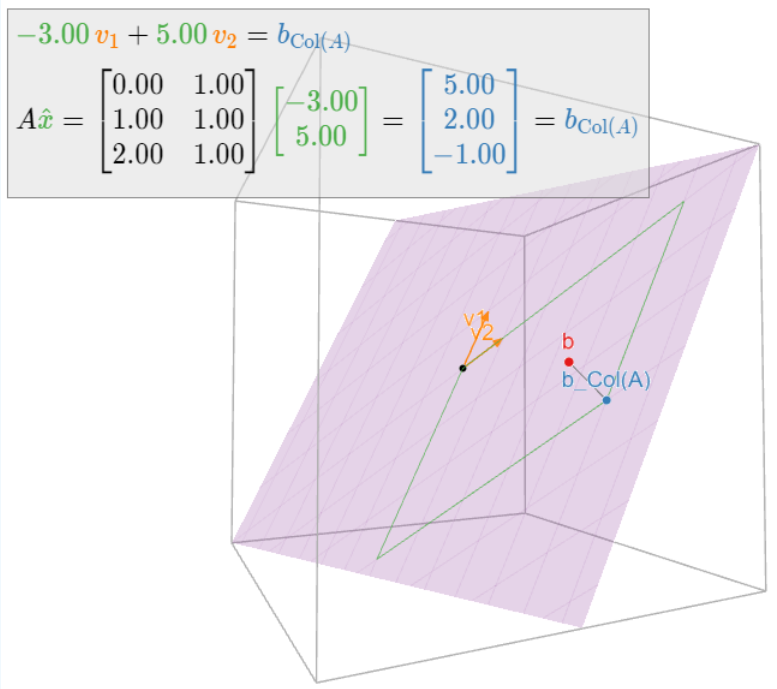


Figure 7.5.5: The violet plane is $\text{Col}(A)$. The closest that Ax can get to b is the closest vector on $\text{Col}(A)$ to b, which is the orthogonal projection $b_{\text{Col}(A)}$ (in blue). The vectors v_1, v_2 are the columns of A, and the coefficients of $\hat{x}$ are the $\mathcal{B}$-coordinates of $b_{\text{Col}(A)}$, where $\mathcal{B} = \{v_1, v_2\}$. lengths of the green lines.

Example 7.5.2

Find the least-squares solutions of $Ax = b$ where:

$$A = \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} \qquad b = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Solution

We have

$$A^TA = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 5 & -1 \\ -1 & 5 \end{pmatrix}$$

and

$$A^Tb = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}.$$

We form an augmented matrix and row reduce:

$$\left(\begin{array}{cc|c} 5 & -1 & 2 \\ -1 & 5 & -2 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & 1/3 \\ 0 & 1 & -1/3 \end{array}\right).$$

Therefore, the only least-squares solution is $\hat{x} = \frac{1}{3}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

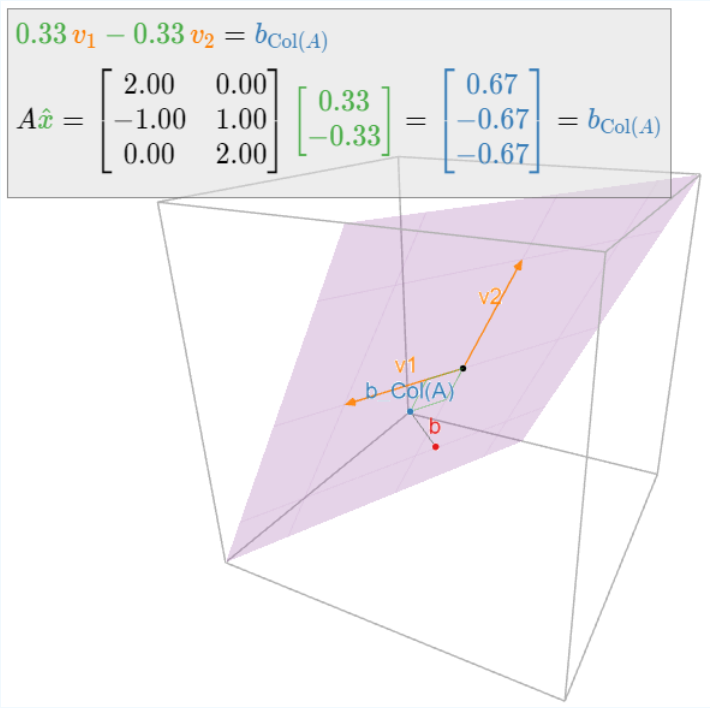


Figure 7.5.6: The red plane is $\mathrm{Col}(A)$. The closest that Ax can get to b is the closest vector on $\mathrm{Col}(A)$ to b, which is the orthogonal projection $b_{\mathrm{Col}(A)}$ (in blue). The vectors v_1, v_2 are the columns of A, and the coefficients of $\hat{x}$ are the $\mathcal{B}$-coordinates of $b_{\mathrm{Col}(A)}$, where $\mathcal{B} = \{v_1, v_2\}$. lengths of the green lines.

The reader may have noticed that we have been careful to say "the least-squares solutions" in the plural, and "a least-squares solution" using the indefinite article. This is because a least-squares solution need not be unique: indeed, if the columns of A are linearly dependent, then $Ax = b_{\mathrm{Col}(A)}$ has infinitely many solutions. The following theorem, which gives equivalent criteria for uniqueness, is an analogue of Corollary 6.3.1 in Section 6.3.

Theorem 7.5.2

Let A be an $m \times n$ matrix and let b be a vector in $\mathbb{R}^m$. The following are equivalent:

1. $Ax = b$ has a unique least-squares solution.
2. The columns of A are linearly independent.
3. A^TA is invertible.

In this case, the least-squares solution is

$$\hat{x} = (A^TA)^{-1}A^Tb.$$

Proof

The set of least-squares solutions of $Ax = b$ is the solution set of the consistent equation $A^TAx = A^Tb$, which is a translate of the solution set of the homogeneous equation $A^TAx = 0$. Since A^TA is a square matrix, the equivalence of 1 and 3 follows from Theorem 5.1.1 in Section 5.1. The set of least squares-solutions is also the solution set of the consistent equation $Ax = b_{\text{Col}(A)}$, which has a unique solution if and only if the columns of A are linearly independent by Recipe: Checking Linear Independence in Section 2.5.

Example 7.5.3: Infinitely many least-squares solutions

Find the least-squares solutions of $Ax = b$ where:

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 1 & 2 & -3 \end{pmatrix} \qquad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

Solution

We have

$$A^TA = \begin{pmatrix} 3 & 3 & -3 \\ 3 & 5 & -7 \\ -3 & -7 & 11 \end{pmatrix} \qquad A^Tb = \begin{pmatrix} 6 \\ 0 \\ 6 \end{pmatrix}.$$

We form an augmented matrix and row reduce:

$$\left(\begin{array}{ccc|c} 3 & 3 & -3 & 6 \\ 3 & 5 & -7 & 0 \\ -3 & -7 & 11 & 6 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 5 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

The free variable is x_3, so the solution set is

$$\begin{cases} x_1 = -x_3 + 5 \\ x_2 = 2x_3 - 3 \\ x_3 = x_3 \end{cases} \xrightarrow[\text{vector form}]{\text{parametric}} \hat{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 5 \\ -3 \\ 0 \end{pmatrix}.$$

For example, taking $x_3 = 0$ and $x_3 = 1$ gives the least-squares solutions

$$\hat{x} = \begin{pmatrix} 5 \\ -3 \\ 0 \end{pmatrix} \quad \text{and} \quad \hat{x} = \begin{pmatrix} 4 \\ -1 \\ 1 \end{pmatrix}.$$

Geometrically, we see that the columns v_1, v_2, v_3 of A are coplanar:

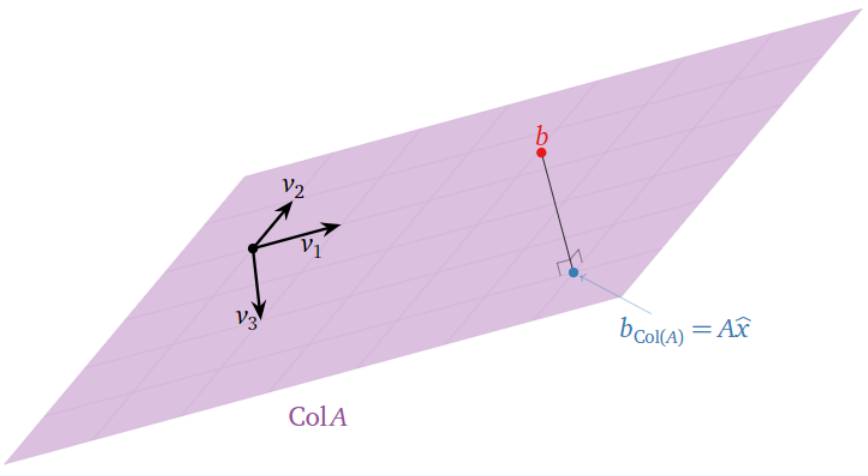


Figure 7.5.7

Therefore, there are many ways of writing $b_{\mathrm{Col}(A)}$ as a linear combination of v_1, v_2, v_3.

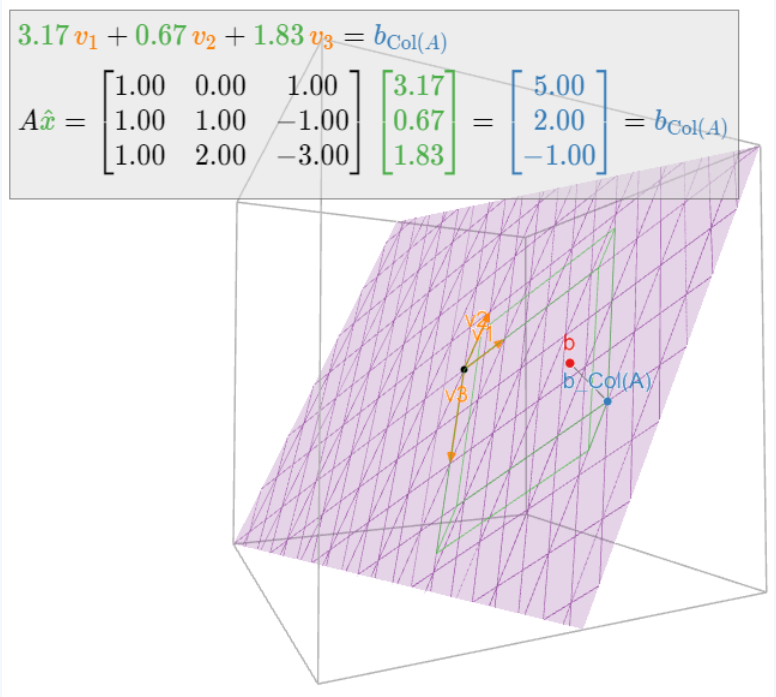


Figure 7.5.8: The three columns of A are coplanar, so there are many least-squares solutions. (The demo picks one solution when you move b.)

As usual, calculations involving projections become easier in the presence of an orthogonal set. Indeed, if A is an $m \times n$ matrix with *orthogonal* columns $u_1, u_2, \ldots, u_m$, then we can use the Theorem 6.4.1 in Section 6.4 to write

$$b_{\mathrm{Col}(A)} = \frac{b \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{b \cdot u_2}{u_2 \cdot u_2} u_2 + \cdots + \frac{b \cdot u_m}{u_m \cdot u_m} u_m = A \begin{pmatrix} (b \cdot u_1)/(u_1 \cdot u_1) \\ (b \cdot u_2)/(u_2 \cdot u_2) \\ \vdots \\ (b \cdot u_m)/(u_m \cdot u_m) \end{pmatrix}.$$

Note that the least-squares solution is unique in this case, since an orthogonal set is linearly independent, Fact 6.4.1 in Section 6.4.

Recipe 2: Compute a Least-Squares Solution

Let A be an $m \times n$ matrix with *orthogonal* columns $u_1, u_2, \ldots, u_m$, and let b be a vector in $\mathbb{R}^n$. Then the least-squares solution of $Ax = b$ is the vector

$$\hat{x} = \left(\frac{b \cdot u_1}{u_1 \cdot u_1}, \frac{b \cdot u_2}{u_2 \cdot u_2}, \ldots, \frac{b \cdot u_m}{u_m \cdot u_m} \right).$$

This formula is particularly useful in the sciences, as matrices with orthogonal columns often arise in nature.

Example 7.5.4

Find the least-squares solution of $Ax = b$ where:

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 4 \end{pmatrix}.$$

Solution

Let u_1, u_2, u_3 be the columns of A. These form an orthogonal set, so

$$\hat{x} = \left(\frac{b \cdot u_1}{u_1 \cdot u_1}, \frac{b \cdot u_2}{u_2 \cdot u_2}, \frac{b \cdot u_3}{u_3 \cdot u_3} \right) = \left(\frac{-3}{2}, \frac{-3}{2}, \frac{8}{4} \right) = \left(-\frac{3}{2}, -\frac{3}{2}, 2 \right).$$

Compare Example 6.4.8 in Section 6.4.

Best-Fit Problems

In this subsection we give an application of the method of least squares to data modeling. We begin with a basic example.

Example 7.5.5: Best-Fit Line

Suppose that we have measured three data points

$$(0, 6), \quad (1, 0), \quad (2, 0),$$

and that our model for these data asserts that the points should lie on a line. Of course, these three points do not actually lie on a single line, but this could be due to errors in our measurement. How do we predict which line they are supposed to lie on?

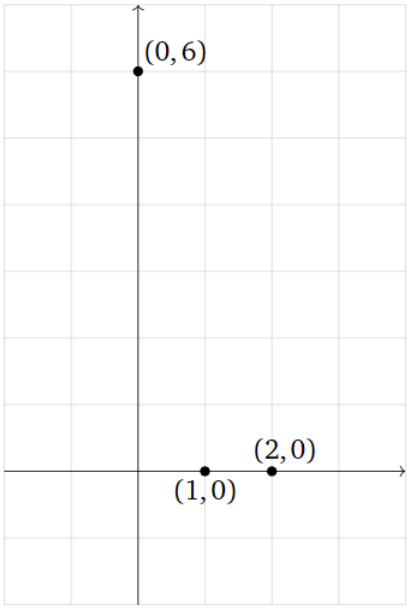


Figure 7.5.9

The general equation for a (non-vertical) line is

$$y = Mx + B.$$

If our three data points were to lie on this line, then the following equations would be satisfied:

$$\begin{aligned} 6 &= M \cdot 0 + B \\ 0 &= M \cdot 1 + B \\ 0 &= M \cdot 2 + B. \end{aligned} \tag{7.5.1}$$

In order to find the best-fit line, we try to solve the above equations in the unknowns M and B. As the three points do not actually lie on a line, there is no actual solution, so instead we compute a least-squares solution.

Putting our linear equations into matrix form, we are trying to solve $Ax = b$ for

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} \qquad x = \begin{pmatrix} M \\ B \end{pmatrix} \qquad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

We solved this least-squares problem in Example 7.5.1: the only least-squares solution to $Ax = b$ is $\hat{x} = \binom{M}{B} = \binom{-3}{5}$, so the best-fit line is

$$y = -3x + 5.$$

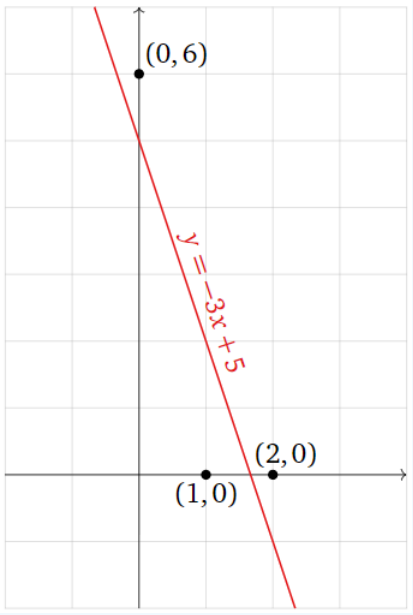


Figure 7.5.10

What exactly is the line $y = f(x) = -3x + 5$ minimizing? The least-squares solution $\hat{x}$ minimizes the sum of the squares of the entries of the vector $b - A\hat{x}$. The vector b is the left-hand side of (7.5.1), and

$$A\begin{pmatrix} -3 \\ 5 \end{pmatrix} = \begin{pmatrix} -3(0)+5 \\ -3(1)+5 \\ -3(2)+5 \end{pmatrix} = \begin{pmatrix} f(0) \\ f(1) \\ f(2) \end{pmatrix}.$$

In other words, $A\hat{x}$ is the vector whose entries are the y-coordinates of the graph of the line at the values of x we specified in our data points, and b is the vector whose entries are the y-coordinates of those data points. The difference $b - A\hat{x}$ is the vertical distance of the graph from the data points:

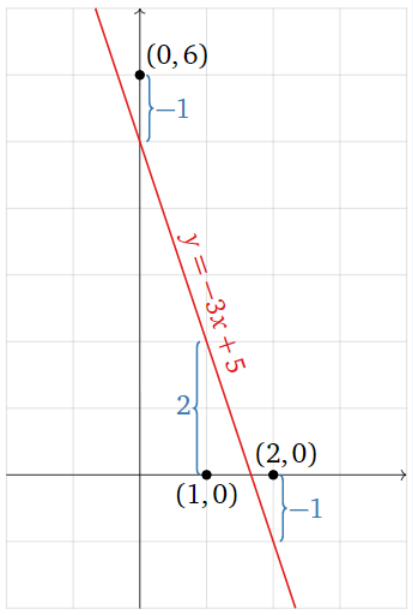


Figure 7.5.11

$$b - A\hat{x} = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} - A \begin{pmatrix} -3 \\ 5 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$$

The best-fit line minimizes the sum of the squares of these vertical distances.

Example 7.5.6: Interactive: Best-fit line

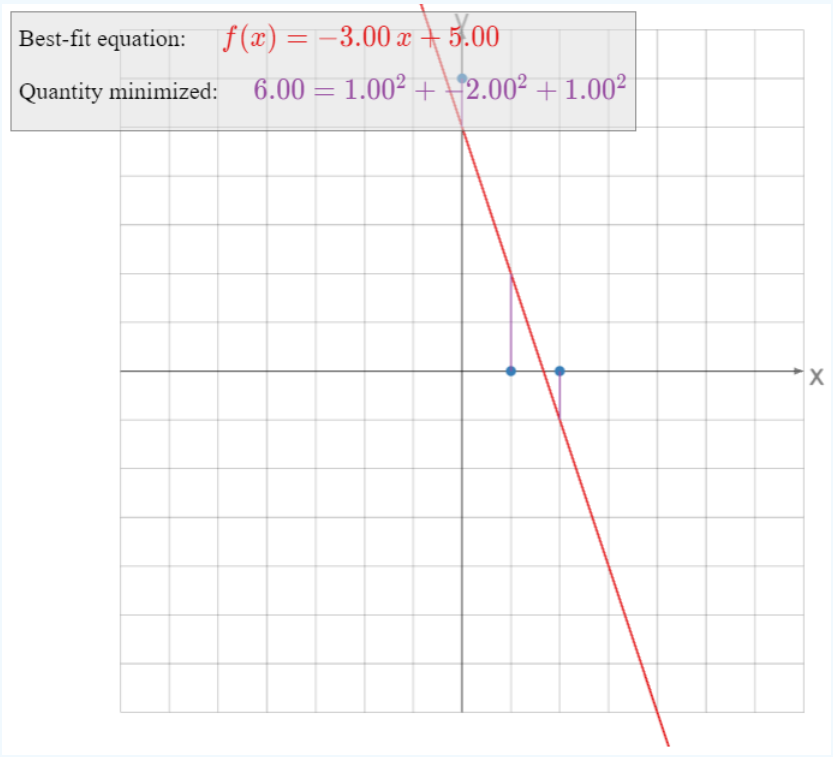


Figure 7.5.12: The best-fit line minimizes the sum of the squares of the vertical distances (violet). Click and drag the points to see how the best-fit line changes.

Example 7.5.7: Best-fit parabola

Find the parabola that best approximates the data points

$$(-1,\ 1/2), \quad (1,\ -1), \quad (2,\ -1/2), \quad (3,\ 2).$$

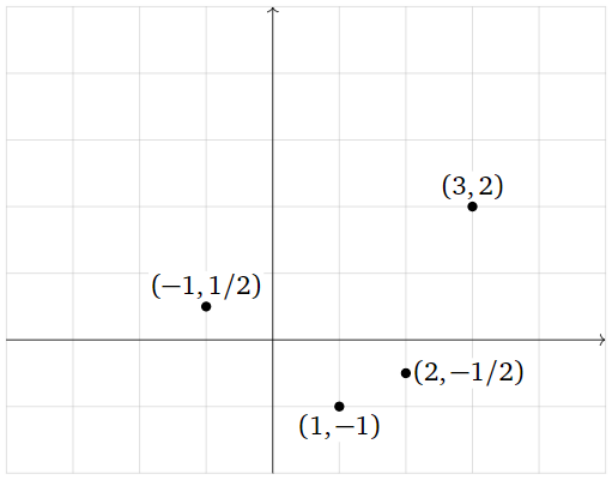


Figure 7.5.13

What quantity is being minimized?

Solution

The general equation for a parabola is

$$y = Bx^2 + Cx + D.$$

If the four points were to lie on this parabola, then the following equations would be satisfied:

$$\begin{aligned} \frac{1}{2} &= B(-1)^2 + C(-1) + D \\ -1 &= B(1)^2 + C(1) + D \\ -\frac{1}{2} &= B(2)^2 + C(2) + D \\ 2 &= B(3)^2 + C(3) + D. \end{aligned} \tag{7.5.2}$$

We treat this as a system of equations in the unknowns B, C, D. In matrix form, we can write this as $Ax = b$ for

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} \qquad x = \begin{pmatrix} B \\ C \\ D \end{pmatrix} \qquad b = \begin{pmatrix} 1/2 \\ -1 \\ -1/2 \\ 2 \end{pmatrix}.$$

We find a least-squares solution by multiplying both sides by the transpose:

$$A^TA = \begin{pmatrix} 99 & 35 & 15 \\ 35 & 15 & 5 \\ 15 & 5 & 4 \end{pmatrix} \qquad A^Tb = \begin{pmatrix} 31/2 \\ 7/2 \\ 1 \end{pmatrix},$$

then forming an augmented matrix and row reducing:

$$\left(\begin{array}{ccc|c} 99 & 35 & 15 & 31/2 \\ 35 & 15 & 5 & 7/2 \\ 15 & 5 & 4 & 1 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 53/88 \\ 0 & 1 & 0 & -379/440 \\ 0 & 0 & 1 & -41/44 \end{array}\right) \implies \hat{x} = \begin{pmatrix} 53/88 \\ -379/440 \\ -41/44 \end{pmatrix}.$$

The best-fit parabola is

$$y = \frac{53}{88}x^2 - \frac{379}{440}x - \frac{41}{44}.$$

Multiplying through by 88, we can write this as

$$88y = 53x^2 - \frac{379}{5}x - 82.$$

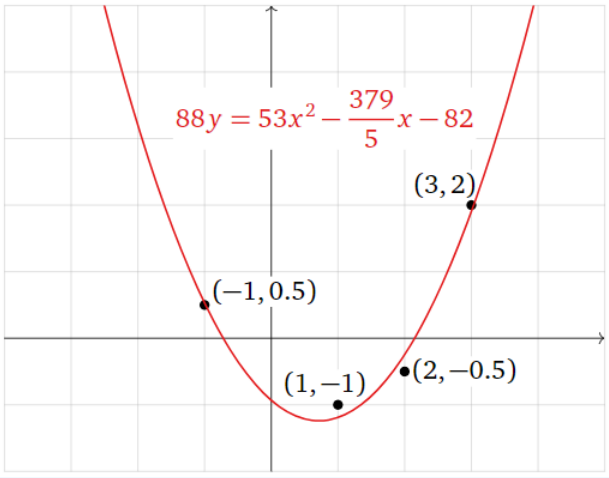


Figure 7.5.14

Now we consider what exactly the parabola $y = f(x)$ is minimizing. The least-squares solution $\hat{x}$ minimizes the sum of the squares of the entries of the vector $b - A\hat{x}$. The vector b is the left-hand side of (7.5.2), and

$$A\hat{x} = \begin{pmatrix} \frac{53}{88}(-1)^2 - \frac{379}{440}(-1) - \frac{41}{44} \\ \frac{53}{88}(1)^2 - \frac{379}{440}(1) - \frac{41}{44} \\ \frac{53}{88}(2)^2 - \frac{379}{440}(2) - \frac{41}{44} \\ \frac{53}{88}(3)^2 - \frac{379}{440}(3) - \frac{41}{44} \end{pmatrix} = \begin{pmatrix} f(-1) \\ f(1) \\ f(2) \\ f(3) \end{pmatrix}.$$

In other words, $A\hat{x}$ is the vector whose entries are the y-coordinates of the graph of the parabola at the values of x we specified in our data points, and b is the vector whose entries are the y-coordinates of those data points. The difference $b - A\hat{x}$ is the vertical distance of the graph from the data points:

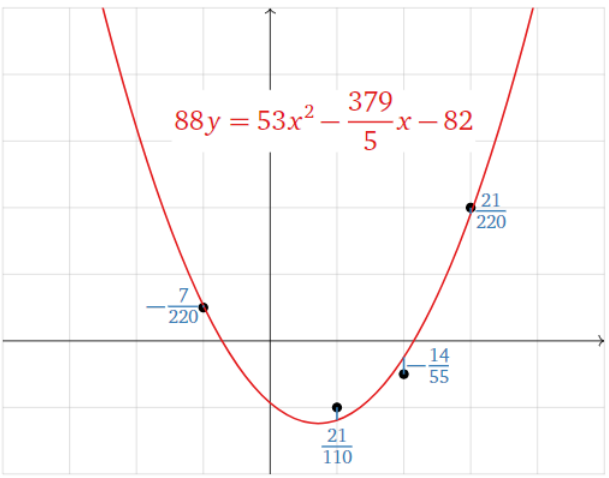


Figure 7.5.15

$$b - A\hat{x} = \begin{pmatrix} 1/2 \\ -1 \\ -1/2 \\ 2 \end{pmatrix} - A \begin{pmatrix} 53/88 \\ -379/440 \\ -41/44 \end{pmatrix} = \begin{pmatrix} -7/220 \\ 21/110 \\ -14/55 \\ 21/220 \end{pmatrix}$$

The best-fit parabola minimizes the sum of the squares of these vertical distances.

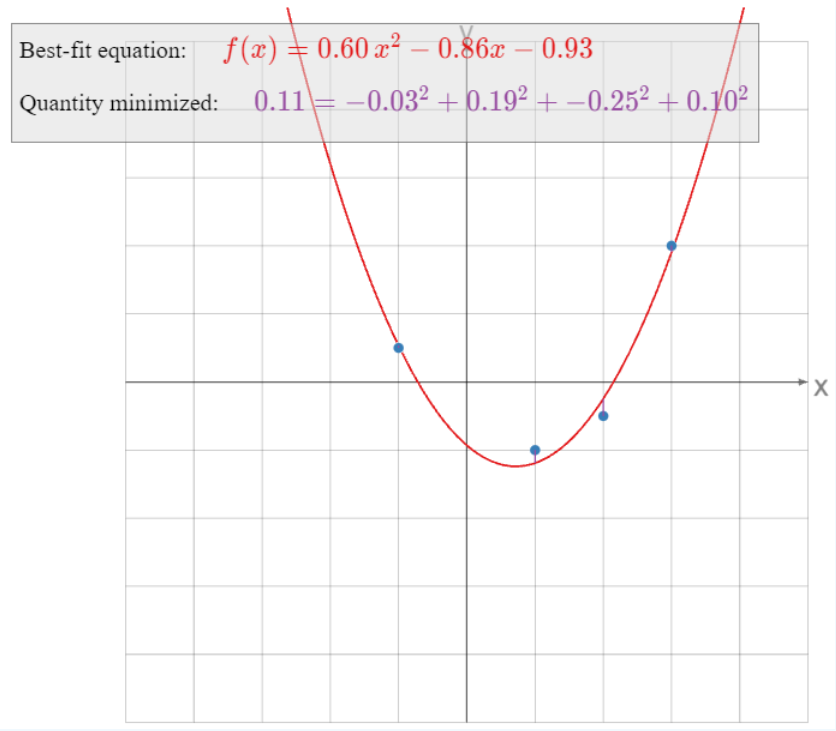


Figure 7.5.16: The best-fit parabola minimizes the sum of the squares of the vertical distances (violet). Click and drag the points to see how the best-fit parabola changes.

Example 7.5.8: Best-fit linear function

Find the linear function $f(x, y)$ that best approximates the following data:

x	y	$f(x,y)$
1	0	0
0	1	1
−1	0	3
0	−1	4

What quantity is being minimized?

Solution

The general equation for a linear function in two variables is

$$f(x, y) = Bx + Cy + D.$$

We want to solve the following system of equations in the unknowns B, C, D:

$$\begin{aligned} B(1) + C(0) + D &= 0 \\ B(0) + C(1) + D &= 1 \\ B(-1) + C(0) + D &= 3 \\ B(0) + C(-1) + D &= 4 \end{aligned} \tag{7.5.3}$$

In matrix form, we can write this as $Ax = b$ for

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \qquad x = \begin{pmatrix} B \\ C \\ D \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 4 \end{pmatrix}.$$

We observe that the columns u_1, u_2, u_3 of A are *orthogonal*, so we can use Recipe 2: Compute a Least-Squares Solution:

$$\hat{x} = \left(\frac{b \cdot u_1}{u_1 \cdot u_1}, \frac{b \cdot u_2}{u_2 \cdot u_2}, \frac{b \cdot u_3}{u_3 \cdot u_3} \right) = \left(\frac{-3}{2}, \frac{-3}{2}, \frac{8}{4} \right) = \left(-\frac{3}{2}, -\frac{3}{2}, 2 \right).$$

We find a least-squares solution by multiplying both sides by the transpose:

$$A^T A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \qquad A^T b = \begin{pmatrix} -3 \\ -3 \\ 8 \end{pmatrix}.$$

The matrix $A^T A$ is diagonal (do you see why that happened?), so it is easy to solve the equation $A^T A x = A^T b$:

$$\begin{pmatrix} 2 & 0 & 0 & -3 \\ 0 & 2 & 0 & -3 \\ 0 & 0 & 4 & 8 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 0 & -3/2 \\ 0 & 1 & 0 & -3/2 \\ 0 & 0 & 1 & 2 \end{pmatrix} \implies \hat{x} = \begin{pmatrix} -3/2 \\ -3/2 \\ 2 \end{pmatrix}.$$

Therefore, the best-fit linear equation is

$$f(x, y) = -\frac{3}{2}x - \frac{3}{2}y + 2.$$

Here is a picture of the graph of $f(x, y)$:

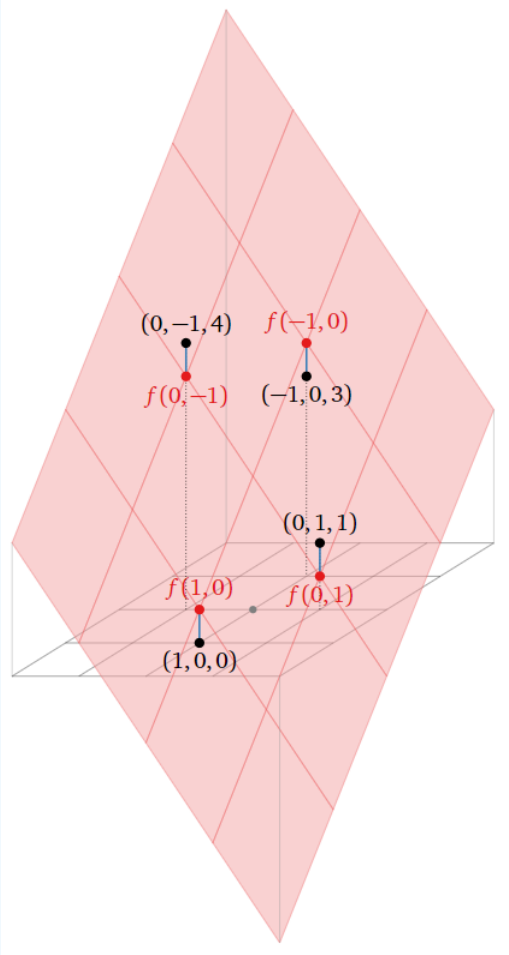


Figure 7.5.17

Now we consider what quantity is being minimized by the function $f(x, y)$. The least-squares solution $\hat{x}$ minimizes the sum of the squares of the entries of the vector $b - A\hat{x}$. The vector b is the right-hand side of (7.5.3), and

$$A\hat{x} = \begin{pmatrix} -\frac{3}{2}(1) & - & \frac{3}{2}(0) & + & 2 \\ -\frac{3}{2}(0) & - & \frac{3}{2}(1) & + & 2 \\ -\frac{3}{2}(-1) & - & \frac{3}{2}(0) & + & 2 \\ -\frac{3}{2}(0) & - & \frac{3}{2}(-1) & + & 2 \end{pmatrix} = \begin{pmatrix} f(1,0) \\ f(0,1) \\ f(-1,0) \\ f(0,-1) \end{pmatrix}.$$

In other words, $A\hat{x}$ is the vector whose entries are the values of f evaluated on the points (x, y) we specified in our data table, and b is the vector whose entries are the desired values of f evaluated at those points. The difference $b - A\hat{x}$ is the vertical distance of the graph from the data points, as indicated in the above picture. The best-fit linear function minimizes the sum of these vertical distances.

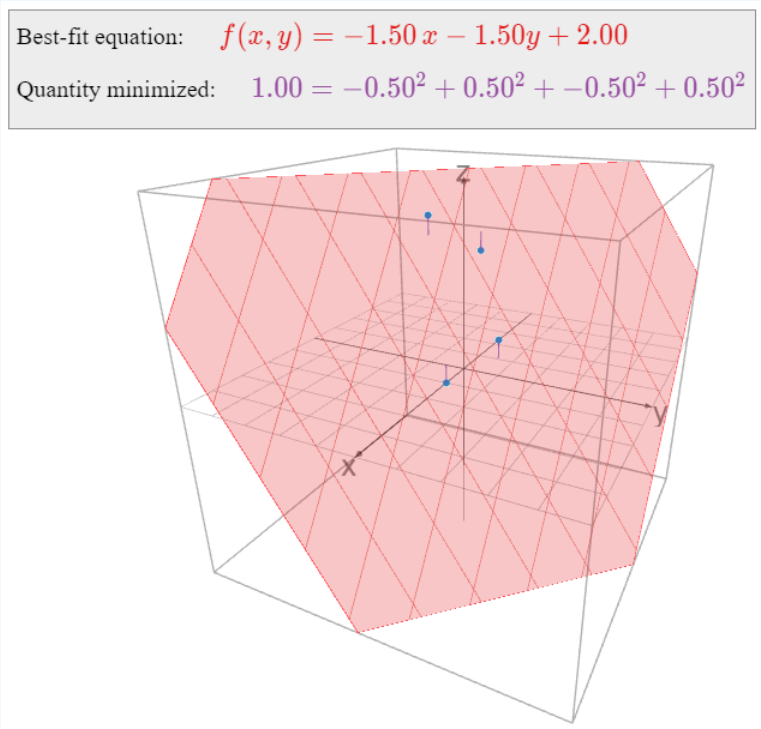


Figure 7.5.18: The best-fit linear function minimizes the sum of the squares of the vertical distances (violet). Click and drag the points to see how the best-fit linear function changes.

All of the above examples have the following form: some number of data points (x, y) are specified, and we want to find a function

$$y = B_1 g_1(x) + B_2 g_2(x) + \cdots + B_m g_m(x)$$

that best approximates these points, where $g_1, g_2, \ldots, g_m$ are fixed functions of x. Indeed, in the best-fit line example we had $g_1(x) = x$ and $g_2(x) = 1$; in the best-fit parabola example we had $g_1(x) = x^2$, $g_2(x) = x$, and $g_3(x) = 1$; and in the best-fit linear function example we had $g_1(x_1, x_2) = x_1$, $g_2(x_1, x_2) = x_2$, and $g_3(x_1, x_2) = 1$ (in this example we take x to be a vector with two entries). We evaluate the above equation on the given data points to obtain a system of linear equations in the unknowns $B_1, B_2, \ldots, B_m$—once we evaluate the g_i, they just become numbers, so it does not matter what they are—and we find the least-squares solution. The resulting best-fit function minimizes the sum of the squares of the vertical distances from the graph of $y = f(x)$ to our original data points.

To emphasize that the nature of the functions g_i really is irrelevant, consider the following example.

Example 7.5.9: Best-fit trigonometric function

What is the best-fit function of the form

$$y = B + C\cos(x) + D\sin(x) + E\cos(2x) + F\sin(2x) + G\cos(3x) + H\sin(3x)$$

passing through the points

$\begin{pmatrix}-4\\-1\end{pmatrix}, \begin{pmatrix}-3\\0\end{pmatrix}, \begin{pmatrix}-2\\-1.5\end{pmatrix}, \begin{pmatrix}-1\\.5\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}, \begin{pmatrix}2\\-.5\end{pmatrix}, \begin{pmatrix}3\\2\end{pmatrix}, \begin{pmatrix}4\\-1\end{pmatrix}$?

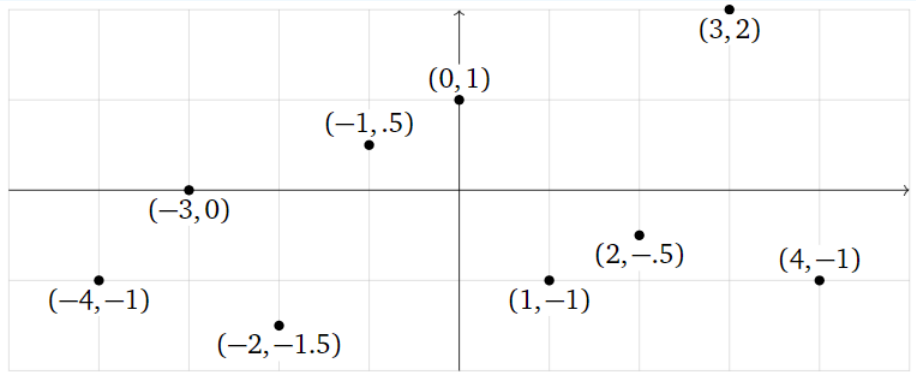


Figure 7.5.19

Solution

We want to solve the system of equations

$$\begin{aligned}
-1 &= B + C\cos(-4) + D\sin(-4) + E\cos(-8) + F\sin(-8) + G\cos(-12) + H\sin(-12)\\
0 &= B + C\cos(-3) + D\sin(-3) + E\cos(-6) + F\sin(-6) + G\cos(-9) + H\sin(-9)\\
-1.5 &= B + C\cos(-2) + D\sin(-2) + E\cos(-4) + F\sin(-4) + G\cos(-6) + H\sin(-6)\\
0.5 &= B + C\cos(-1) + D\sin(-1) + E\cos(-2) + F\sin(-2) + G\cos(-3) + H\sin(-3)\\
1 &= B + C\cos(0) + D\sin(0) + E\cos(0) + F\sin(0) + G\cos(0) + H\sin(0)\\
-1 &= B + C\cos(1) + D\sin(1) + E\cos(2) + F\sin(2) + G\cos(3) + H\sin(3)\\
-0.5 &= B + C\cos(2) + D\sin(2) + E\cos(4) + F\sin(4) + G\cos(6) + H\sin(6)\\
2 &= B + C\cos(3) + D\sin(3) + E\cos(6) + F\sin(6) + G\cos(9) + H\sin(9)\\
-1 &= B + C\cos(4) + D\sin(4) + E\cos(8) + F\sin(8) + G\cos(12) + H\sin(12).
\end{aligned}$$

All of the terms in these equations are *numbers*, except for the unknowns B, C, D, E, F, G, H:

$$\begin{aligned}
-1 &= B - 0.6536C + 0.7568D - 0.1455E - 0.9894F + 0.8439G + 0.5366H\\
0 &= B - 0.9900C - 0.1411D + 0.9602E + 0.2794F - 0.9111G - 0.4121H\\
-1.5 &= B - 0.4161C - 0.9093D - 0.6536E + 0.7568F + 0.9602G + 0.2794H\\
0.5 &= B + 0.5403C - 0.8415D - 0.4161E - 0.9093F - 0.9900G - 0.1411H\\
1 &= B + C + E + G\\
-1 &= B + 0.5403C + 0.8415D - 0.4161E + 0.9093F - 0.9900G + 0.1411H\\
-0.5 &= B - 0.4161C + 0.9093D - 0.6536E - 0.7568F + 0.9602G - 0.2794H\\
2 &= B - 0.9900C + 0.1411D + 0.9602E - 0.2794F - 0.9111G + 0.4121H\\
-1 &= B - 0.6536C - 0.7568D - 0.1455E + 0.9894F + 0.8439G - 0.5366H.
\end{aligned}$$

Hence we want to solve the least-squares problem

$$\begin{pmatrix}
1 & -0.6536 & 0.7568 & -0.1455 & -0.9894 & 0.8439 & 0.5366\\
1 & -0.9900 & -0.1411 & 0.9602 & 0.2794 & -0.9111 & -0.4121\\
1 & -0.4161 & -0.9093 & -0.6536 & 0.7568 & 0.9602 & 0.2794\\
1 & 0.5403 & -0.8415 & -0.4161 & -0.9093 & -0.9900 & -0.1411\\
1 & 1 & 0 & 1 & 0 & 1 & 0\\
1 & 0.5403 & 0.8415 & -0.4161 & 0.9093 & -0.9900 & 0.1411\\
1 & -0.4161 & 0.9093 & -0.6536 & -0.7568 & 0.9602 & -0.2794\\
1 & -0.9900 & 0.1411 & 0.9602 & -0.2794 & -0.9111 & 0.4121\\
1 & -0.6536 & -0.7568 & -0.1455 & 0.9894 & 0.8439 & -0.5366
\end{pmatrix}
\begin{pmatrix}B\\C\\D\\E\\F\\G\\H\end{pmatrix}
=
\begin{pmatrix}-1\\0\\-1.5\\0.5\\1\\-1\\-0.5\\2\\-1\end{pmatrix}.$$

We find the least-squares solution with the aid of a computer:

$$\hat{x} \approx \begin{pmatrix} -0.1435 \\ 0.2611 \\ -0.2337 \\ 1.116 \\ -0.5997 \\ -0.2767 \\ 0.1076 \end{pmatrix}.$$

Therefore, the best-fit function is

$$\begin{aligned} y \ &\approx -0.1435 + 0.2611\cos(x) - 0.2337\sin(x) + 1.116\cos(2x) - 0.5997\sin(2x) \\ &\quad - 0.2767\cos(3x) + 0.1076\sin(3x). \end{aligned}$$

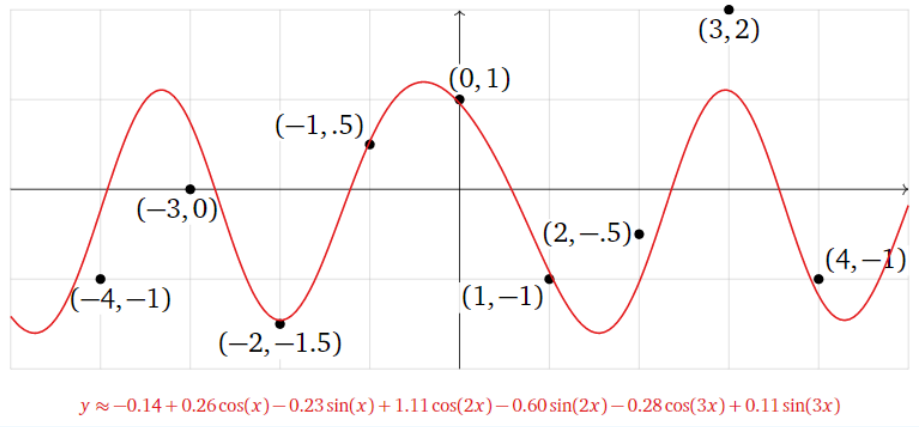


Figure 7.5.20

As in the previous examples, the best-fit function minimizes the sum of the squares of the vertical distances from the graph of $y = f(x)$ to the data points.

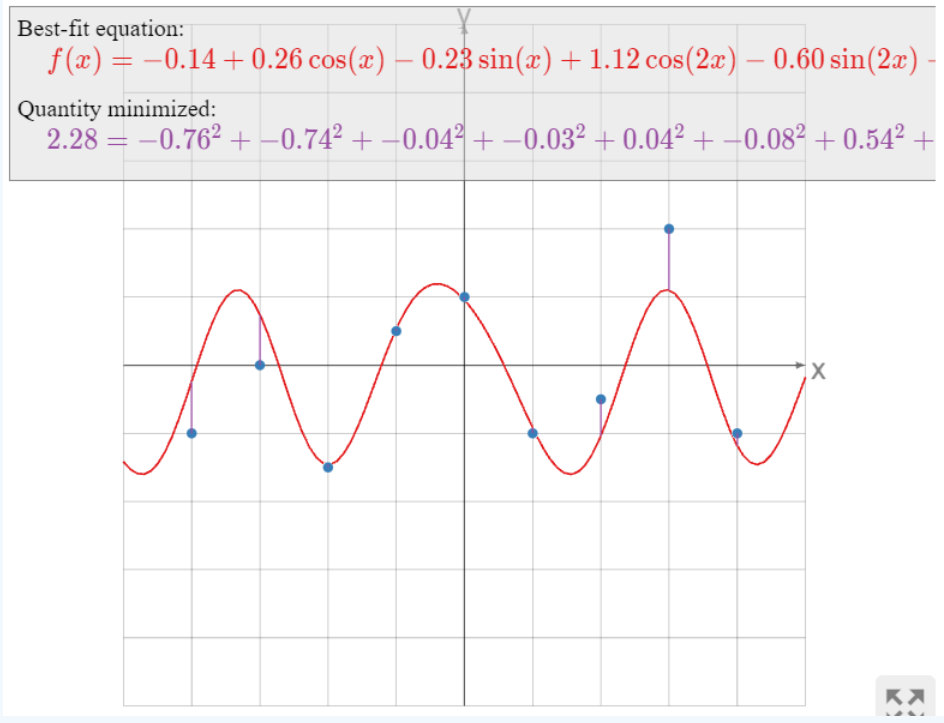


Figure 7.5.21: The best-fit function minimizes the sum of the squares of the vertical distances (violet). Click and drag the points to see how the best-fit function changes.

The next example has a somewhat different flavor from the previous ones.

Example 7.5.10: Best-fit ellipse

Find the best-fit ellipse through the points

$$(0,2),\ (2,1),\ (1,-1),\ (-1,-2),\ (-3,1),\ (-1,-1).$$

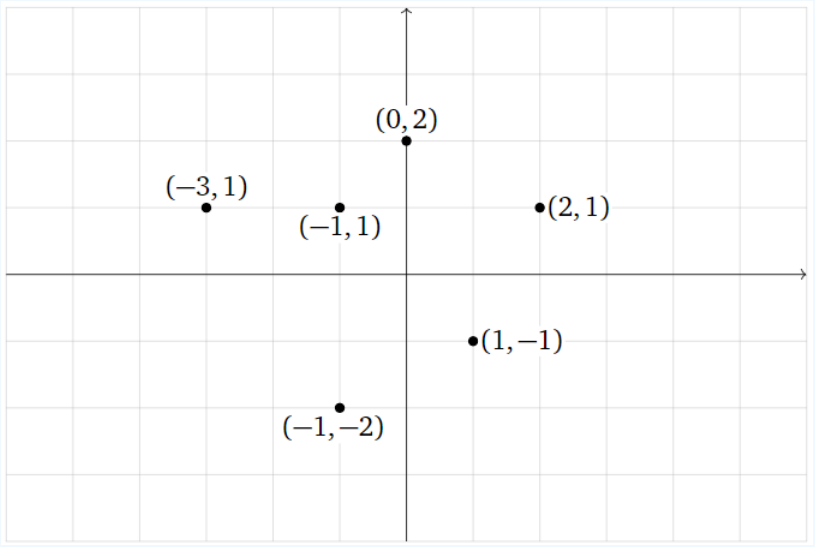


Figure 7.5.22

What quantity is being minimized?

Solution

The general equation for an ellipse (actually, for a nondegenerate conic section) is

$$x^2 + By^2 + Cxy + Dx + Ey + F = 0.$$

This is an *implicit equation*: the ellipse is the set of all solutions of the equation, just like the unit circle is the set of solutions of $x^2 + y^2 = 1$. To say that our data points lie on the ellipse means that the above equation is satisfied for the given values of x and y:

$$\begin{array}{rcrcrcrcrcr} (0)^2 & + & B(2)^2 & + & C(0)(2) & + & D(0) & + & E(2) & + & F & = & 0 \\ (2)^2 & + & B(1)^2 & + & C(2)(1) & + & D(2) & + & E(1) & + & F & = & 0 \\ (1)^2 & + & B(-1)^2 & + & C(1)(-1) & + & D(1) & + & E(-1) & + & F & = & 0 \\ (-1)^2 & + & B(-2)^2 & + & C(-1)(2) & + & D(-1) & + & E(-2) & + & F & = & 0 \\ (-3)^2 & + & B(1)^2 & + & C(-3)(1) & + & D(-3) & + & E(1) & + & F & = & 0 \\ (-1)^2 & + & B(-1)^2 & + & C(-1)(-1) & + & D(-1) & + & E(-1) & + & F & = & 0. \end{array} \tag{7.5.4}$$

To put this in matrix form, we move the constant terms to the right-hand side of the equals sign; then we can write this as $Ax = b$ for

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 \end{pmatrix} \quad x = \begin{pmatrix} B \\ C \\ D \\ E \\ F \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \\ -1 \end{pmatrix}.$$

We compute

$$A^T A = \begin{pmatrix} 36 & 7 & -5 & 0 & 12 \\ 7 & 19 & 9 & -5 & 1 \\ -5 & 9 & 16 & 1 & -2 \\ 0 & -5 & 1 & 12 & 0 \\ 12 & 1 & -2 & 0 & 6 \end{pmatrix} \qquad A^T b = \begin{pmatrix} -19 \\ 17 \\ 20 \\ -9 \\ -16 \end{pmatrix}.$$

We form an augmented matrix and row reduce:

$$\left(\begin{array}{ccccc|c} 36 & 7 & -5 & 0 & 12 & -19 \\ 7 & 19 & 9 & -5 & 1 & 17 \\ -5 & 9 & 16 & 1 & -2 & 20 \\ 0 & -5 & 1 & 12 & 0 & -9 \\ 12 & 1 & -2 & 0 & 6 & -16 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 405/266 \\ 0 & 1 & 0 & 0 & 0 & -89/133 \\ 0 & 0 & 1 & 0 & 0 & 201/133 \\ 0 & 0 & 0 & 1 & 0 & -123/266 \\ 0 & 0 & 0 & 0 & 1 & -687/133 \end{array}\right).$$

The least-squares solution is

$$\hat{x} = \begin{pmatrix} 405/266 \\ -89/133 \\ 201/133 \\ -123/266 \\ -687/133 \end{pmatrix},$$

so the best-fit ellipse is

$$x^2 + \frac{405}{266}y^2 - \frac{89}{133}xy + \frac{201}{133}x - \frac{123}{266}y - \frac{687}{133} = 0.$$

Multiplying through by 266, we can write this as

$$266x^2 + 405y^2 - 178xy + 402x - 123y - 1374 = 0.$$

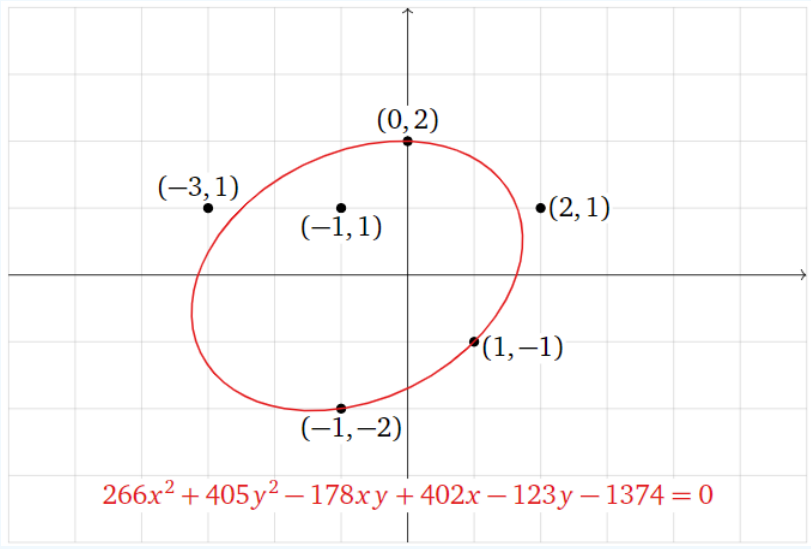


Figure 7.5.23

Now we consider the question of what quantity is minimized by this ellipse. The least-squares solution $\hat{x}$ minimizes the sum of the squares of the entries of the vector $b - A\hat{x}$, or equivalently, of $A\hat{x} - b$. The vector $-b$ contains the constant terms of the left-hand sides of (7.5.4), and

$$A\hat{x} = \begin{pmatrix} \frac{405}{266}(2)^2 - \frac{89}{133}(0)(2) + \frac{201}{133}(0) - \frac{123}{266}(2) - \frac{687}{133} \\ \frac{405}{266}(1)^2 - \frac{89}{133}(2)(1) + \frac{201}{133}(2) - \frac{123}{266}(1) - \frac{687}{133} \\ \frac{405}{266}(-1)^2 - \frac{89}{133}(1)(-1) + \frac{201}{133}(1) - \frac{123}{266}(-1) - \frac{687}{133} \\ \frac{405}{266}(-2)^2 - \frac{89}{133}(-1)(-2) + \frac{201}{133}(-1) - \frac{123}{266}(-2) - \frac{687}{133} \\ \frac{405}{266}(1)^2 - \frac{89}{133}(-3)(1) + \frac{201}{133}(-3) - \frac{123}{266}(1) - \frac{687}{133} \\ \frac{405}{266}(-1)^2 - \frac{89}{133}(-1)(-1) + \frac{201}{133}(-1) - \frac{123}{266}(-1) - \frac{687}{133} \end{pmatrix}$$

contains the rest of the terms on the left-hand side of (7.5.4). Therefore, the entries of $A\hat{x} - b$ are the quantities obtained by evaluating the function

$$f(x,y) = x^2 + \frac{405}{266}y^2 - \frac{89}{133}xy + \frac{201}{133}x - \frac{123}{266}y - \frac{687}{133}$$

on the given data points.

If our data points actually lay on the ellipse defined by $f(x,y) = 0$, then evaluating $f(x,y)$ on our data points would always yield zero, so $A\hat{x} - b$ would be the zero vector. This is not the case; instead, $A\hat{x} - b$ contains the *actual* values of $f(x,y)$ when evaluated on our data points. The quantity being minimized is the sum of the squares of these values:

$$\text{minimized} = f(0,2)^2 + f(2,1)^2 + f(1,-1)^2 + f(-1,-2)^2 + f(-3,1)^2 + f(-1,-1)^2.$$

One way to visualize this is as follows. We can put this best-fit problem into the framework of Example 7.5.8 by asking to find an equation of the form

$$f(x,y) = x^2 + By^2 + Cxy + Dx + Ey + F$$

which best approximates the data table

x	y	$f(x,y)$
0	2	0
2	1	0
1	−1	0
−1	−2	0
−3	1	0
−1	−1	0.

The resulting function minimizes the sum of the squares of the vertical distances from these data points $(0,2,0)$, $(2,1,0)$, $\ldots$, which lie on the xy-plane, to the graph of $f(x,y)$.

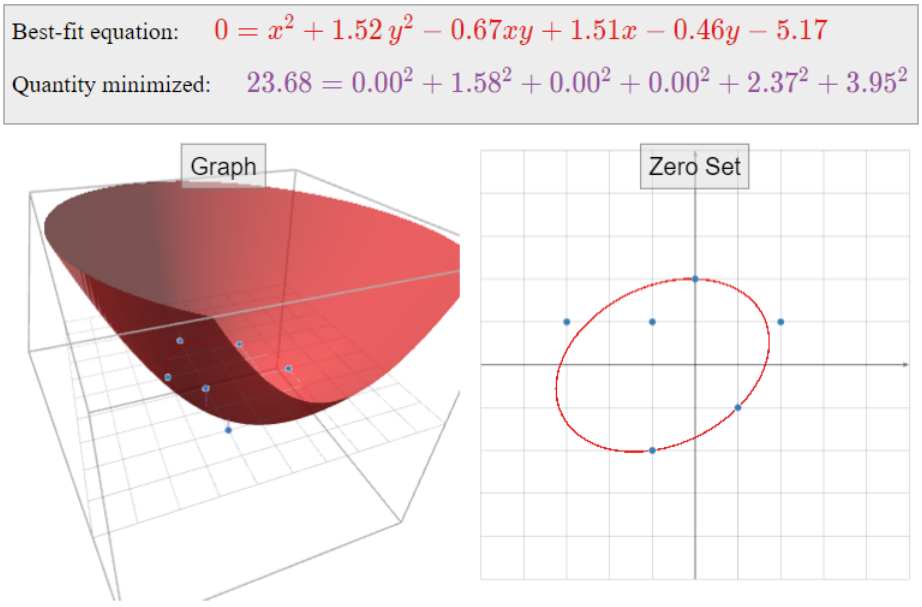


Figure 7.5.24: The best-fit ellipse minimizes the sum of the squares of the vertical distances (violet) from the points $(x, y, 0)$ to the graph of $f(x, y)$ on the left. The ellipse itself is the zero set of $f(x, y)$, on the right. Click and drag the points on the right to see how the best-fit ellipse changes. Can you arrange the points so that the best-fit conic section is actually a hyperbola?

Note 7.5.3

Gauss invented the method of least squares to find a best-fit ellipse: he correctly predicted the (elliptical) orbit of the asteroid Ceres as it passed behind the sun in 1801.

7.5: The Method of Least Squares is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by LibreTexts.

- **6.5: The Method of Least Squares** by Dan Margalit & Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

7.6: Orthogonal Diagonalization

Recall that an $n \times n$ matrix A is diagonalizable if and only if it has n linearly independent eigenvectors. Moreover, the matrix P with these eigenvectors as columns is a diagonalizing matrix for A, that is

$$P^{-1}AP \text{ is diagonal.}$$

As we have seen, the really nice bases of $\mathbb{R}^n$ are the orthogonal ones, so a natural question is: which $n \times n$ matrices have an *orthogonal* basis of eigenvectors? These turn out to be precisely the symmetric matrices, and this is the main result of this section.

Before proceeding, recall that an orthogonal set of vectors is called *orthonormal* if $\|\mathbf{v}\| = 1$ for each vector $\mathbf{v}$ in the set, and that any orthogonal set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ can be "*normalized*", that is converted into an orthonormal set $\{\frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1, \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2, \ldots, \frac{1}{\|\mathbf{v}_k\|}\mathbf{v}_k\}$. In particular, if a matrix A has n orthogonal eigenvectors, they can (by normalizing) be taken to be orthonormal. The corresponding diagonalizing matrix P has orthonormal columns, and such matrices are very easy to invert.

The following conditions are equivalent for an $n \times n$ matrix P.

1. P is invertible and $P^{-1} = P^T$.
2. The rows of P are orthonormal.
3. The columns of P are orthonormal.

First recall that condition (1) is equivalent to $PP^T = I$. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ denote the rows of P. Then $\mathbf{x}_j^T$ is the jth column of P^T, so the (i, j)-entry of PP^T is $\mathbf{x}_i \bullet \mathbf{x}_j$. Thus $PP^T = I$ means that $\mathbf{x}_i \bullet \mathbf{x}_j = 0$ if $i \neq j$ and $\mathbf{x}_i \bullet \mathbf{x}_j = 1$ if $i = j$. Hence condition (1) is equivalent to (2). The proof of the equivalence of (1) and (3) is similar.

Orthogonal Matrices: An $n \times n$ matrix P is called an **orthogonal matrix**if it satisfies one (and hence all) of the conditions in Theorem [thm:024227].

The rotation matrix $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal for any angle θ.

These orthogonal matrices have the virtue that they are easy to invert—simply take the transpose. But they have many other important properties as well. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator, we will prove that T is distance preserving if and only if its matrix is orthogonal. In particular, the matrices of rotations and reflections about the origin in $\mathbb{R}^2$ and $\mathbb{R}^3$ are all orthogonal.

It is not enough that the rows of a matrix A are merely orthogonal for A to be an orthogonal matrix. Here is an example.

The matrix $\begin{bmatrix} 2 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}$ has orthogonal rows but the columns are not orthogonal. However, if the rows are normalized, the resulting matrix $\begin{bmatrix} \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ is orthogonal (so the columns are now orthonormal as the reader can verify).

If P and Q are orthogonal matrices, then PQ is also orthogonal, as is $P^{-1} = P^T$.

P and Q are invertible, so PQ is also invertible and

$$(PQ)^{-1} = Q^{-1}P^{-1} = Q^TP^T = (PQ)^T$$

Hence PQ is orthogonal. Similarly,

$$(P^{-1})^{-1} = P = (P^T)^T = (P^{-1})^T$$

shows that P^{-1} is orthogonal.

Orthogonally Diagonalizable Matrices: An $n \times n$ matrix A is said to be **orthogonally diagonalizable** when an orthogonal matrix P can be found such that $P^{-1}AP = P^TAP$ is diagonal.

This condition turns out to characterize the symmetric matrices.

Principal Axes Theorem The following conditions are equivalent for an $n \times n$ matrix A.

1. A has an orthonormal set of n eigenvectors.
2. A is orthogonally diagonalizable.
3. A is symmetric.

(1) $\Leftrightarrow$ (2). Given (1), let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be orthonormal eigenvectors of A. Then $P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \end{bmatrix}$ is orthogonal, and $P^{-1}AP$ is diagonal. This proves (2). Conversely, given (2) let $P^{-1}AP$ be diagonal where P is orthogonal. If $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are the columns of P then $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is an orthonormal basis of $\mathbb{R}^n$ that consists of eigenvectors of A. This proves (1).

(2) $\Rightarrow$ (3). If $P^TAP = D$ is diagonal, where $P^{-1} = P^T$, then $A = PDP^T$. But $D^T = D$, so this gives $A^T = P^{TT}D^TP^T = PDP^T = A$.

(3) $\Rightarrow$ (2). If A is an $n \times n$ symmetric matrix, we proceed by induction on n. If $n = 1$, A is already diagonal. If $n > 1$, assume that (3) $\Rightarrow$ (2) for $(n-1) \times (n-1)$ symmetric matrices. By Theorem [thm:016397] let λ_1 be a (real) eigenvalue of A, and let $A\mathbf{x}_1 = \lambda_1\mathbf{x}_1$, where $\|\mathbf{x}_1\| = 1$. Use the Gram-Schmidt algorithm to find an orthonormal basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ for $\mathbb{R}^n$. Let $P_1 = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \end{bmatrix}$, so P_1 is an orthogonal matrix and $P_1^TAP_1 = \begin{bmatrix} \lambda_1 & B \\ 0 & A_1 \end{bmatrix}$ in block form by Lemma [lem:016161]. But $P_1^TAP_1$ is symmetric (A is), so it follows that $B = 0$ and A_1 is symmetric. Then, by induction, there exists an $(n-1) \times (n-1)$ orthogonal matrix Q such that $Q^TA_1Q = D_1$ is diagonal. Observe that $P_2 = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$ is orthogonal, and compute:

$$\begin{aligned} (P_1P_2)^TA(P_1P_2) &= P_2^T(P_1^TAP_1)P_2 \\ &= \begin{bmatrix} 1 & 0 \\ 0 & Q^T \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & 0 \\ 0 & D_1 \end{bmatrix} \end{aligned}$$

is diagonal. Because P_1P_2 is orthogonal, this proves (2).

A set of orthonormal eigenvectors of a symmetric matrix A is called a set of **principal axes** for A. The name comes from geometry, and this is discussed in Section [sec:8_8]. Because the eigenvalues of a (real) symmetric matrix are real, Theorem [thm:024303] is also called the **real spectral theorem**, and the set of distinct eigenvalues is called the **spectrum** of the matrix. In full generality, the spectral theorem is a similar result for matrices with complex entries.

Find an orthogonal matrix P such that $P^{-1}AP$ is diagonal, where $A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ -1 & 2 & 5 \end{bmatrix}$.

The characteristic polynomial of A is (adding twice row 1 to row 2):

$$c_A(x) = \det \begin{bmatrix} x-1 & 0 & 1 \\ 0 & x-1 & -2 \\ 1 & -2 & x-5 \end{bmatrix} = x(x-1)(x-6)$$

Thus the eigenvalues are $\lambda = 0$, 1, and 6, and corresponding eigenvectors are

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}$$

respectively. Moreover, by what appears to be remarkably good luck, these eigenvectors are *orthogonal*. We have $\|\mathbf{x}_1\|^2 = 6$, $\|\mathbf{x}_2\|^2 = 5$, and $\|\mathbf{x}_3\|^2 = 30$, so

$$P = \begin{bmatrix} \frac{1}{\sqrt{6}}\mathbf{x}_1 & \frac{1}{\sqrt{5}}\mathbf{x}_2 & \frac{1}{\sqrt{30}}\mathbf{x}_3 \end{bmatrix} = \frac{1}{\sqrt{30}} \begin{bmatrix} \sqrt{5} & 2\sqrt{6} & -1 \\ -2\sqrt{5} & \sqrt{6} & 2 \\ \sqrt{5} & 0 & 5 \end{bmatrix}$$

is an orthogonal matrix. Thus $P^{-1} = P^T$ and

$$P^TAP = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

by the diagonalization algorithm.

Actually, the fact that the eigenvectors in Example are orthogonal is no coincidence. Theorem guarantees they are linearly independent (they correspond to distinct eigenvalues); the fact that the matrix is *symmetric* implies that they are orthogonal. To prove this we need the following useful fact about symmetric matrices.

If A is an $n \times n$ symmetric matrix, then

$$(A\mathbf{x}) \bullet \mathbf{y} = \mathbf{x} \bullet (A\mathbf{y})$$

for all columns $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.

Recall that $\mathbf{x} \bullet \mathbf{y} = \mathbf{x}^T\mathbf{y}$ for all columns $\mathbf{x}$ and $\mathbf{y}$. Because $A^T = A$, we get

$$(A\mathbf{x}) \bullet \mathbf{y} = (A\mathbf{x})^T\mathbf{y} = \mathbf{x}^TA^T\mathbf{y} = \mathbf{x}^TA\mathbf{y} = \mathbf{x} \bullet (A\mathbf{y})$$

If A is a symmetric matrix, then eigenvectors of A corresponding to distinct eigenvalues are orthogonal.

Let $A\mathbf{x} = \lambda\mathbf{x}$ and $A\mathbf{y} = \mu\mathbf{y}$, where $\lambda \neq \mu$. Using Theorem [thm:024396], we compute

$$\lambda(\mathbf{x} \bullet \mathbf{y}) = (\lambda\mathbf{x}) \bullet \mathbf{y} = (A\mathbf{x}) \bullet \mathbf{y} = \mathbf{x} \bullet (A\mathbf{y}) = \mathbf{x} \bullet (\mu\mathbf{y}) = \mu(\mathbf{x} \bullet \mathbf{y})$$

Hence $(\lambda - \mu)(\mathbf{x} \bullet \mathbf{y}) = 0$, and so $\mathbf{x} \bullet \mathbf{y} = 0$ because $\lambda \neq \mu$.

Now the procedure for diagonalizing a symmetric $n \times n$ matrix is clear. Find the distinct eigenvalues (all real by Theorem [thm:016397]) and find orthonormal bases for each eigenspace (the Gram-Schmidt algorithm may be needed). Then the set of all these basis vectors is orthonormal (by Theorem [thm:024407]) and contains n vectors. Here is an example.

Orthogonally diagonalize the symmetric matrix $A = \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}$.

The characteristic polynomial is

$$c_A(x) = \det \begin{bmatrix} x-8 & 2 & -2 \\ 2 & x-5 & -4 \\ -2 & -4 & x-5 \end{bmatrix} = x(x-9)^2$$

Hence the distinct eigenvalues are 0 and 9 of multiplicities 1 and 2, respectively, so $dim\ (E_0) = 1$ and $dim\ (E_9) = 2$ by Theorem [thm:016250] (A is diagonalizable, being symmetric). Gaussian elimination gives

$$E_0(A) = span\ \{\mathbf{x}_1\}, \enskip \mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}, \quad \text{and} \quad E_9(A) = span\ \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

The eigenvectors in E_9 are both orthogonal to $\mathbf{x}_1$ as Theorem [thm:024407] guarantees, but not to each other. However, the Gram-Schmidt process yields an orthogonal basis

$$\{\mathbf{x}_2, \mathbf{x}_3\} \text{ of } E_9(A) \quad \text{where} \quad \mathbf{x}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \text{ and } \mathbf{x}_3 = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$$

Normalizing gives orthonormal vectors $\{\frac{1}{3}\mathbf{x}_1, \frac{1}{\sqrt{5}}\mathbf{x}_2, \frac{1}{3\sqrt{5}}\mathbf{x}_3\}$, so

$$P = \begin{bmatrix} \frac{1}{3}\mathbf{x}_1 & \frac{1}{\sqrt{5}}\mathbf{x}_2 & \frac{1}{3\sqrt{5}}\mathbf{x}_3 \end{bmatrix} = \frac{1}{3\sqrt{5}} \begin{bmatrix} \sqrt{5} & -6 & 2 \\ 2\sqrt{5} & 3 & 4 \\ -2\sqrt{5} & 0 & 5 \end{bmatrix}$$

is an orthogonal matrix such that $P^{-1}AP$ is diagonal.

It is worth noting that other, more convenient, diagonalizing matrices P exist. For example, $\mathbf{y}_2 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$ and $\mathbf{y}_3 = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}$ lie in $E_9(A)$ and they are orthogonal. Moreover, they both have norm 3 (as does $\mathbf{x}_1$), so

$$Q = \begin{bmatrix} \frac{1}{3}\mathbf{x}_1 & \frac{1}{3}\mathbf{y}_2 & \frac{1}{3}\mathbf{y}_3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{bmatrix}$$

is a nicer orthogonal matrix with the property that $Q^{-1}AQ$ is diagonal.

If A is symmetric and a set of orthogonal eigenvectors of A is given, the eigenvectors are called principal axes of A. The name comes from geometry. An expression $q = ax_1^2 + bx_1x_2 + cx_2^2$ is called a **quadratic form** in the variables x_1 and x_2, and the graph of the equation $q = 1$ is called a **conic** in these variables. For example, if $q = x_1x_2$, the graph of $q = 1$ is given in the first diagram.

But if we introduce new variables y_1 and y_2 by setting $x_1 = y_1 + y_2$ and $x_2 = y_1 - y_2$, then q becomes $q = y_1^2 - y_2^2$, a diagonal form with no cross term y_1y_2 (see the second diagram). Because of this, the y_1 and y_2 axes are called the principal axes for the conic (hence the name). Orthogonal diagonalization provides a systematic method for finding principal axes. Here is an illustration.

Find principal axes for the quadratic form $q = x_1^2 - 4x_1x_2 + x_2^2$.

In order to utilize diagonalization, we first express q in matrix form. Observe that

$$q = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

The matrix here is not symmetric, but we can remedy that by writing

$$q = x_1^2 - 2x_1x_2 - 2x_2x_1 + x_2^2$$

Then we have

$$q = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{x}^TA\mathbf{x}$$

where $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix}$ is symmetric. The eigenvalues of A are $\lambda_1 = 3$ and $\lambda_2 = -1$, with corresponding (orthogonal) eigenvectors $\mathbf{x}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Since $\|\mathbf{x}_1\| = \|\mathbf{x}_2\| = \sqrt{2}$, so

$$P = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \text{ is orthogonal and } P^TAP = D = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}$$

Now define new variables $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{y}$ by $\mathbf{y} = P^T\mathbf{x}$, equivalently $\mathbf{x} = P\mathbf{y}$ (since $P^{-1} = P^T$). Hence

$$y_1 = \frac{1}{\sqrt{2}}(x_1 - x_2) \quad \text{and} \quad y_2 = \frac{1}{\sqrt{2}}(x_1 + x_2)$$

In terms of y_1 and y_2, q takes the form

$$q = \mathbf{x}^TA\mathbf{x} = (P\mathbf{y})^TA(P\mathbf{y}) = \mathbf{y}^T(P^TAP)\mathbf{y} = \mathbf{y}^TD\mathbf{y} = 3y_1^2 - y_2^2$$

Note that $\mathbf{y} = P^T\mathbf{x}$ is obtained from $\mathbf{x}$ by a counterclockwise rotation of $\frac{\pi}{4}$ (see Theorem [thm:004693]).

Observe that the quadratic form q in Example [exa:024463] can be diagonalized in other ways. For example

$$q = x_1^2 - 4x_1x_2 + x_2^2 = z_1^2 - \frac{1}{3}z_2^2$$

where $z_1 = x_1 - 2x_2$ and $z_2 = 3x_2$. We examine this more carefully in Section [sec:8_8].

If we are willing to replace “diagonal” by “upper triangular” in the principal axes theorem, we can weaken the requirement that A is symmetric to insisting only that A has real eigenvalues.

Triangulation Theorem: If A is an $n \times n$ matrix with n real eigenvalues, an orthogonal matrix P exists such that $P^T AP$ is upper triangular.

We modify the proof of Theorem [thm:024303]. If $A\mathbf{x}_1 = \lambda_1 \mathbf{x}_1$ where $\|\mathbf{x}_1\| = 1$, let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be an orthonormal basis of $\mathbb{R}^n$, and let $P_1 = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$. Then P_1 is orthogonal and $P_1^T A P_1 = \begin{bmatrix} \lambda_1 & B \\ 0 & A_1 \end{bmatrix}$ in block form. By induction, let $Q^T A_1 Q = T_1$ be upper triangular where Q is of size $(n-1) \times (n-1)$ and orthogonal. Then $P_2 = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$ is orthogonal, so $P = P_1 P_2$ is also orthogonal and $P^T AP = \begin{bmatrix} \lambda_1 & BQ \\ 0 & T_1 \end{bmatrix}$ is upper triangular.

The proof of Theorem gives no way to construct the matrix P. However, an algorithm will be given in Section [sec:11_1] where an improved version of Theorem [thm:024503] is presented. In a different direction, a version of Theorem [thm:024503] holds for an arbitrary matrix with complex entries (Schur’s theorem).

As for a diagonal matrix, the eigenvalues of an upper triangular matrix are displayed along the main diagonal. Because A and $P^T AP$ have the same determinant and trace whenever P is orthogonal, Theorem [thm:024503] gives:

If A is an $n \times n$ matrix with real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (possibly not all distinct), then $\det A = \lambda_1 \lambda_2 \ldots \lambda_n$ and $\func{tr} A = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

This corollary remains true even if the eigenvalues are not real (using Schur’s theorem).

This page titled 7.6: Orthogonal Diagonalization is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by W. Keith Nicholson (Lyryx Learning Inc.) via source content that was edited to the style and standards of the LibreTexts platform.

- **8.2: Orthogonal Diagonalization** by W. Keith Nicholson is licensed CC BY-NC-SA 4.0. Original source: https://lyryx.com/linear-algebra-applications.

CHAPTER OVERVIEW

8: Some Prerequisite Topics

The topics presented in this section are important concepts in mathematics and therefore should be examined.

8.1: Sets and Set Notation
8.2: Well Ordering and Mathematical Induction

This page titled 8: Some Prerequisite Topics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

8.1: 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 $\emptyset$, which 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 8.1.1: 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 8.1.2: 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 8.1.3: 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.

This page titled 8.1: Sets and Set Notation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **10.1: Sets and Set Notation** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

8.2: Well Ordering and Mathematical 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 8.2.1: 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

$$\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1s} \\ a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rs} \end{array}$$

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 8.2.1: 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 8.2.1: 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 8.2.1: 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.

Mathematical induction is a very useful device for proving theorems about the integers. The procedure is as follows.

Procedure 8.2.1: 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 8.2.1: Proving by Induction

Prove by induction that $\sum_{k=1}^{n} k^2 = \dfrac{n(n+1)(2n+1)}{6}$.

Solution

By Procedure 8.2.1, 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.

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} \tag{8.2.1}$$

We want to show that this is true for $n+1$.

Suppose we add $(n+1)^2$ to both sides of Equation (8.2.1).

$$\begin{aligned} \sum_{k=1}^{n+1} k^2 &= \sum_{k=1}^{n} k^2 + (n+1)^2 \\ &= \frac{n(n+1)(2n+1)}{6} + (n+1)^2 \end{aligned}$$

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 8.2.2: 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 8.2.1 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.

This page titled 8.2: Well Ordering and Mathematical Induction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Ken Kuttler (Lyryx) via source content that was edited to the style and standards of the LibreTexts platform.

- **10.2: Well Ordering and Induction** by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

Index

A

augmented matrix
- 1.2: Gaussian Elimination

B

basic variables
- 1.3: Uniqueness of the Reduced Row-Echelon Form

block diagonalization
- 6.4: Complex Eigenvalues

C

characteristic polynomial
- 6.2: The Characteristic Polynomial

closest point
- 7.1: Dot Products and Orthogonality

cofactor
- 3.1: Basic Techniques

cofactor matrix
- 3.4: Applications of the Determinant

complex eigenvalues
- 6.4: Complex Eigenvalues

constant matrix
- 1.2: Gaussian Elimination

Coordinate Isomorphism
- 5.11: The Matrix of a Linear Transformation

D

damping factor
- 6.5: Stochastic Matrices

determinant
- 3.1: Basic Techniques

determinants
- 3.2: Properties of Determinants

diagonalization
- 6.3: Diagonalization

difference equations
- 6.5: Stochastic Matrices

dot product
- 7.1: Dot Products and Orthogonality

E

eigenspace
- 6.1: Eigenvalues and Eigenvectors

eigenvalue
- 6.1: Eigenvalues and Eigenvectors

eigenvector
- 6.1: Eigenvalues and Eigenvectors

elementary matrix
- 2.8: Elementary Matrices

elementary row operations
- 1.2: Gaussian Elimination

elements
- 8.1: Sets and Set Notation

equivalent matrices
- 1.3: Uniqueness of the Reduced Row-Echelon Form

F

free variables
- 1.3: Uniqueness of the Reduced Row-Echelon Form

G

Gaussian elimination
- 1.2: Gaussian Elimination

general solution
- 4.9: The General Solution of a Linear System

I

identity matrix
- 2.6: The Identity and Inverses

importance matrix
- 6.5: Stochastic Matrices

importance rule
- 6.5: Stochastic Matrices

Induction
- 8.2: Well Ordering and Mathematical Induction

interpolating polynomial
- 3.4: Applications of the Determinant

isomorphism
- 4.6: Isomorphisms
- 5.9: A Isomorphisms

K

kernel
- 4.7: The Kernel and Image of A Linear Map
- 4.9: The General Solution of a Linear System
- 5.10: The Kernel and Image of a Linear Map

Kronecker delta
- 2.6: The Identity and Inverses

L

Laplace Expansion
- 3.1: Basic Techniques

linear map
- 4.6: Isomorphisms

linear transformation
- 4.1: Linear Transformations
- 4.6: Isomorphisms
- 4.9: The General Solution of a Linear System

LU factorization
- 2.10: LU Factorization

M

matrix
- 2.1: Matrix Arithmetic

matrix inverse
- 2.7: Finding the Inverse of a Matrix
- 2.9: More on Matrix Inverses

matrix multiplication
- 2.4: Properties of Matrix Multiplication

matrix transformations
- 4.1: Linear Transformations

minor of a matrix
- 3.1: Basic Techniques

minors or a matrix
- 3.1: Basic Techniques

N

norm of vector
- 7.1: Dot Products and Orthogonality

null space
- 4.9: The General Solution of a Linear System

O

onto transformation
- 5.9: A Isomorphisms

orthogonal
- 7.1: Dot Products and Orthogonality

orthogonal complement
- 7.2: Orthogonal Complements

orthogonal decomposition
- 7.3: Orthogonal Projection

orthogonal diagonalization
- 7.6: Orthogonal Diagonalization

orthogonal projection
- 7.3: Orthogonal Projection

P

PageRank
- 6.5: Stochastic Matrices

Parallelogram law
- 1.4: Vectors

particular solution
- 4.9: The General Solution of a Linear System

Perron–Frobenius theorem
- 6.5: Stochastic Matrices

point
- 1.4: Vectors

R

random surfer interpretation
- 6.5: Stochastic Matrices

Rank
- 6.5: Stochastic Matrices

row operations
- 3.2: Properties of Determinants

row space
- 7.2: Orthogonal Complements

S

set
- 8.1: Sets and Set Notation

set notation
- 8.1: Sets and Set Notation

similar matrices
- 6.6: Similarity

skew lines
- 1.1: Systems of Linear Equations

skew symmetric matrix
- 2.5: The Transpose

square matrix
- 2.1: Matrix Arithmetic

stochastic matrices
- 6.5: Stochastic Matrices

T

transforms
- 4.1: Linear Transformations

transpose
- 2.5: The Transpose

triangular matrix
- 3.1: Basic Techniques

U
union
8.1: Sets and Set Notation
unit vector
7.1: Dot Products and Orthogonality
V
vector space
5.1: Vector Spaces
vectors
1.4: Vectors
W
Well Ordering Principle
8.2: Well Ordering and Mathematical Induction
Z
zero matrix
2.1: Matrix Arithmetic

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: MAT 04C Linear Algebra (Kravets)

Webpages: 81

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY 4.0: 58% (47 pages)
- GNU Free Documentation License: 27.2% (22 pages)
- Undeclared: 13.6% (11 pages)
- CC BY-NC-SA 4.0: 1.2% (1 page)

By Page

- MAT 04C Linear Algebra (Kravets) - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Systems of Linear Equations - *CC BY 4.0*
    - 1.1: Systems of Linear Equations - *CC BY 4.0*
    - 1.2: Gaussian Elimination - *CC BY 4.0*
    - 1.3: Uniqueness of the Reduced Row-Echelon Form - *CC BY 4.0*
    - 1.4: Vectors - *GNU Free Documentation License*
    - 1.5: Vector Equations and Spans - *GNU Free Documentation License*
    - 1.6: Matrix Equations - *GNU Free Documentation License*
    - 1.7: Solution Sets - *GNU Free Documentation License*
    - 1.8: Applications of Linear Systems - *CC BY 4.0*
    - 1.9: Linear Independence - *GNU Free Documentation License*
    - 1.10: Exercises - *CC BY 4.0*
  - 2: Matrices - *CC BY 4.0*
    - 2.1: Matrix Arithmetic - *CC BY 4.0*
    - 2.2: Multiplication of Matrices - *CC BY 4.0*
    - 2.3: The ijth Entry of a Product - *CC BY 4.0*
    - 2.4: Properties of Matrix Multiplication - *CC BY 4.0*
    - 2.5: The Transpose - *CC BY 4.0*
    - 2.6: The Identity and Inverses - *CC BY 4.0*
    - 2.7: Finding the Inverse of a Matrix - *CC BY 4.0*
    - 2.8: Elementary Matrices - *CC BY 4.0*
    - 2.9: More on Matrix Inverses - *CC BY 4.0*
    - 2.10: LU Factorization - *CC BY 4.0*
    - 2.E: Exercises - *CC BY 4.0*
  - 3: Determinants - *CC BY 4.0*
    - 3.1: Basic Techniques - *CC BY 4.0*
    - 3.2: Properties of Determinants - *CC BY 4.0*
    - 3.3: Finding Determinants using Row Operations - *CC BY 4.0*
    - 3.4: Applications of the Determinant - *CC BY 4.0*
    - 3.E: Exercises - *CC BY 4.0*
  - 4: Linear Transformations - *CC BY 4.0*
    - 4.1: Linear Transformations - *CC BY 4.0*
    - 4.2: The Matrix of a Linear Transformation I - *CC BY 4.0*
    - 4.3: Properties of Linear Transformations. Invertible Linear Transformations. - *CC BY 4.0*
    - 4.4: Special Linear Transformations and Applications to Computer Graphics - *CC BY 4.0*
    - 4.5: One-to-One and Onto Transformations - *CC BY 4.0*
    - 4.6: Isomorphisms - *CC BY 4.0*
    - 4.7: The Kernel and Image of A Linear Map - *CC BY 4.0*
    - 4.8: The Matrix of a Linear Transformation II - *CC BY 4.0*
    - 4.9: The General Solution of a Linear System - *CC BY 4.0*
    - 4.E: Exercises - *CC BY 4.0*
  - 5: Vector Spaces and Subspaces - *CC BY 4.0*
    - 5.1: Vector Spaces - *CC BY 4.0*
    - 5.2: Subspaces - *GNU Free Documentation License*
    - 5.3: Basis and Dimension - *GNU Free Documentation License*
    - 5.4: Rank and Homogeneous Systems - *CC BY 4.0*
    - 5.5: The Rank Theorem - *GNU Free Documentation License*
    - 5.6: Bases as Coordinate Systems - *GNU Free Documentation License*
    - 5.7: Spanning, Linear Independence and Basis in R^n - *CC BY 4.0*

- 5.8: A Linear Transformations - *CC BY 4.0*
- 5.9: A Isomorphisms - *CC BY 4.0*
- 5.10: The Kernel and Image of a Linear Map - *CC BY 4.0*
- 5.11: The Matrix of a Linear Transformation - *CC BY 4.0*
- 5.E: Exercises - *CC BY 4.0*

6: Eigenvalues and Eigenvectors - *GNU Free Documentation License*

- 6.1: Eigenvalues and Eigenvectors - *GNU Free Documentation License*
- 6.2: The Characteristic Polynomial - *GNU Free Documentation License*
- 6.3: Diagonalization - *GNU Free Documentation License*
- 6.4: Complex Eigenvalues - *GNU Free Documentation License*
- 6.5: Stochastic Matrices - *GNU Free Documentation License*
- 6.6: Similarity - *GNU Free Documentation License*

7: Orthogonality - *GNU Free Documentation License*

- 7.1: Dot Products and Orthogonality - *GNU Free Documentation License*
- 7.2: Orthogonal Complements - *GNU Free Documentation License*
- 7.3: Orthogonal Projection - *GNU Free Documentation License*
- 7.4: Orthogonal Sets and Gram-Schmidt Process - *GNU Free Documentation License*
- 7.5: The Method of Least Squares - *GNU Free Documentation License*
- 7.6: Orthogonal Diagonalization - *CC BY-NC-SA 4.0*

8: Some Prerequisite Topics - *CC BY 4.0*

- 8.1: Sets and Set Notation - *CC BY 4.0*
- 8.2: Well Ordering and Mathematical Induction - *CC BY 4.0*

Back Matter - *Undeclared*

- Index - *Undeclared*
- Glossary - *Undeclared*
- Detailed Licensing - *Undeclared*
- Detailed Licensing - *Undeclared*

Detailed Licensing

Overview

Title: MAT 04C Linear Algebra (Kravets)

Webpages: 81

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY 4.0: 58% (47 pages)
- GNU Free Documentation License: 27.2% (22 pages)
- Undeclared: 13.6% (11 pages)
- CC BY-NC-SA 4.0: 1.2% (1 page)

By Page

- MAT 04C Linear Algebra (Kravets) - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Systems of Linear Equations - *CC BY 4.0*
    - 1.1: Systems of Linear Equations - *CC BY 4.0*
    - 1.2: Gaussian Elimination - *CC BY 4.0*
    - 1.3: Uniqueness of the Reduced Row-Echelon Form - *CC BY 4.0*
    - 1.4: Vectors - *GNU Free Documentation License*
    - 1.5: Vector Equations and Spans - *GNU Free Documentation License*
    - 1.6: Matrix Equations - *GNU Free Documentation License*
    - 1.7: Solution Sets - *GNU Free Documentation License*
    - 1.8: Applications of Linear Systems - *CC BY 4.0*
    - 1.9: Linear Independence - *GNU Free Documentation License*
    - 1.10: Exercises - *CC BY 4.0*
  - 2: Matrices - *CC BY 4.0*
    - 2.1: Matrix Arithmetic - *CC BY 4.0*
    - 2.2: Multiplication of Matrices - *CC BY 4.0*
    - 2.3: The ijth Entry of a Product - *CC BY 4.0*
    - 2.4: Properties of Matrix Multiplication - *CC BY 4.0*
    - 2.5: The Transpose - *CC BY 4.0*
    - 2.6: The Identity and Inverses - *CC BY 4.0*
    - 2.7: Finding the Inverse of a Matrix - *CC BY 4.0*
    - 2.8: Elementary Matrices - *CC BY 4.0*
    - 2.9: More on Matrix Inverses - *CC BY 4.0*
    - 2.10: LU Factorization - *CC BY 4.0*
    - 2.E: Exercises - *CC BY 4.0*
  - 3: Determinants - *CC BY 4.0*
    - 3.1: Basic Techniques - *CC BY 4.0*
    - 3.2: Properties of Determinants - *CC BY 4.0*
    - 3.3: Finding Determinants using Row Operations - *CC BY 4.0*
    - 3.4: Applications of the Determinant - *CC BY 4.0*
    - 3.E: Exercises - *CC BY 4.0*
  - 4: Linear Transformations - *CC BY 4.0*
    - 4.1: Linear Transformations - *CC BY 4.0*
    - 4.2: The Matrix of a Linear Transformation I - *CC BY 4.0*
    - 4.3: Properties of Linear Transformations. Invertible Linear Transformations. - *CC BY 4.0*
    - 4.4: Special Linear Transformations and Applications to Computer Graphics - *CC BY 4.0*
    - 4.5: One-to-One and Onto Transformations - *CC BY 4.0*
    - 4.6: Isomorphisms - *CC BY 4.0*
    - 4.7: The Kernel and Image of A Linear Map - *CC BY 4.0*
    - 4.8: The Matrix of a Linear Transformation II - *CC BY 4.0*
    - 4.9: The General Solution of a Linear System - *CC BY 4.0*
    - 4.E: Exercises - *CC BY 4.0*
  - 5: Vector Spaces and Subspaces - *CC BY 4.0*
    - 5.1: Vector Spaces - *CC BY 4.0*
    - 5.2: Subspaces - *GNU Free Documentation License*
    - 5.3: Basis and Dimension - *GNU Free Documentation License*
    - 5.4: Rank and Homogeneous Systems - *CC BY 4.0*
    - 5.5: The Rank Theorem - *GNU Free Documentation License*
    - 5.6: Bases as Coordinate Systems - *GNU Free Documentation License*
    - 5.7: Spanning, Linear Independence and Basis in R^n - *CC BY 4.0*

- 5.8: A Linear Transformations - *CC BY 4.0*
- 5.9: A Isomorphisms - *CC BY 4.0*
- 5.10: The Kernel and Image of a Linear Map - *CC BY 4.0*
- 5.11: The Matrix of a Linear Transformation - *CC BY 4.0*
- 5.E: Exercises - *CC BY 4.0*

6: Eigenvalues and Eigenvectors - *GNU Free Documentation License*

- 6.1: Eigenvalues and Eigenvectors - *GNU Free Documentation License*
- 6.2: The Characteristic Polynomial - *GNU Free Documentation License*
- 6.3: Diagonalization - *GNU Free Documentation License*
- 6.4: Complex Eigenvalues - *GNU Free Documentation License*
- 6.5: Stochastic Matrices - *GNU Free Documentation License*
- 6.6: Similarity - *GNU Free Documentation License*

7: Orthogonality - *GNU Free Documentation License*

- 7.1: Dot Products and Orthogonality - *GNU Free Documentation License*
- 7.2: Orthogonal Complements - *GNU Free Documentation License*
- 7.3: Orthogonal Projection - *GNU Free Documentation License*
- 7.4: Orthogonal Sets and Gram-Schmidt Process - *GNU Free Documentation License*
- 7.5: The Method of Least Squares - *GNU Free Documentation License*
- 7.6: Orthogonal Diagonalization - *CC BY-NC-SA 4.0*

8: Some Prerequisite Topics - *CC BY 4.0*

- 8.1: Sets and Set Notation - *CC BY 4.0*
- 8.2: Well Ordering and Mathematical Induction - *CC BY 4.0*

Back Matter - *Undeclared*

- Index - *Undeclared*
- Glossary - *Undeclared*
- Detailed Licensing - *Undeclared*
- Detailed Licensing - *Undeclared*