LINEAR ALGEBRA AND ITS APPLICATION



Amira Alkeswani
Cañada College

LibreTexts™

Linear Algebra and Its Application

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 an 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 borne 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 were 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.

The LibreTexts name, logo, and book covers are trademarked by LibreTexts, Inc. (a 501(c)(3) not-for-profit organization) and protected against unauthorized use.

This text was compiled on 08/30/2026

TABLE OF CONTENTS

Licensing

0: Why Linear Algebra?

1: Solving Systems of Linear Equations

1.1: Systems of Linear Equations

1.2: Matrix Notation and Row Reduction

1.3: Rank and Homogeneous Systems

1.4: Chemistry Reaction-Balancing Chemical Reaction

1.5: Applications to Physics and Resistor Networks

1.6: Chapters Exercises

2: Matrices

2.1: Matrix Addition, Scalar Multiplication, and Transposition

2.2: Matrix Multiplication

2.3: The Identity and Inverses

2.4: Finding the Inverse of a Matrix

2.5: Elementary Matrices

2.6: LU Factorization

3: Determinants

3.1: Basic Techniques

3.1E: Exercises for Section 3.1

3.2: Properties of Determinants

3.2E: Exercises for Section 3.2

3.3: Application of the Determinant to Inverses; Cramer's Rule

3.3E: Exercises for Section 3.3

3.4: Determinants and Geometry

3.4E: Exercises for Section 3.4

4: Vector Spaces - R^n

4.1: Subspaces

4.2: Review of Vectors

4.2E: Exercises for Section 4.1

4.3: Dot and Cross Product

4.3E: Exercises for Section 4.2

4.4: Lines and Planes

4.4E: Exercises for Section 4.3

4.5: Spanning Sets in R^n

4.5.E: Exercise for Section 4.4

4.6: Linear Independence
4.6.E: Exercise for Section 4.5
4.7: Subspaces
4.8: Subspaces and Bases
4.8.E: Exercise for Section 4.6
4.9: Row, Column and Null Spaces
4.9.E: Exercise for Section 4.7
4.10: Dot Products and Orthogonality
4.11: Orthogonal Vectors and Matrices
4.11.E: Exercise for Section 4.8
4.12: Gram-Schmidt Process
4.12.E: Exercises for Section 4.9
4.13: Orthogonal Complements
4.14: Orthogonal Projections
4.14.E: Exercise for Section 4.10
4.15: Orthogonal Projection
4.16: Least Squares Approximation
4.16.E: Exercises for Section 4.11
4.17: Vector Spaces and Subspace
4.18: Amira Vector Space Axioms
4.19: Amira Vector Subspace of R^n
4.20: The Real Vector Space ^R^n
4.E: Exercises

5: Linear Transformations
5.1: Prelude to Linear Transformations and Matrix Algebra
5.2: Linear Transformations
5.3: Matrix Transformations
5.4: One-to-one and Onto Transformations
5.5: The Invertible Matrix Theorem

6: Eigenvalues and Eigenvectors
6.1: An introduction to Eigenvalues and Eigenvectors
6.2: Finding eigenvalues and eigenvectors
6.3: Diagonalization, similarity, and powers of a matrix
6.4: Dynamical Systems
6.5: Markov chains and Google's PageRank algorithm

7: Honors Projects
7.1: Stochastic Matrices
7.2: An Application to Markov Chains

7.2.1: Absorbing Markov Chains
7.2.1.1: Absorbing Markov Chains (Exercises)
7.2.2: Introduction to Markov Chains
7.2.2.1: Introduction to Markov Chains (Exercises)
7.2.3: Regular Markov Chains
7.2.3.1: Regular Markov Chains (Exercises)
7.3: The Method of Least Squares
7.4: An Application to Linear Recurrences
7.4E: An Application to Linear Recurrences Exercises
7.5: An Application to Systems of Differential Equations
7.5E: An Application to Systems of Differential Equations Exercises
7.6: An Application to Computer Graphics
7.6E: An Application to Computer Graphics Exercises
7.7: An Application to Input-Output Economic Models
7.7E: An Application to Input-Output Economic Models Exercises
7.8: The Complex Number System
7.8.1: Solve Systems of Linear Equations with Complex Coefficients
Index
Glossary
Detailed Licensing
Detailed Licensing
Detailed Licensing

This page titled Linear Algebra and Its Application is shared under a license and was authored, remixed, and/or curated by Amira Alkeswani, Cañada College.

Licensing

This page titled Licensing is shared under a license and was authored, remixed, and/or curated by LibreTexts.

0: Why Linear Algebra?

Why It Matters

Linear algebra is the study of vectors, matrices, and linear transformations—a framework for understanding relationships between quantities. It is foundational to both pure and applied mathematics, providing tools to model and solve complex problems efficiently. By mastering linear algebra, you gain the ability to think critically about structure, patterns, and systems that appear across disciplines.

Real-World Applications

Linear algebra is everywhere in science and technology:

- **Engineering & Physics:** Analyze forces, electrical circuits, and mechanical systems.
- **Computer Science:** Power graphics, machine learning algorithms, and data compression.
- **Economics & Social Sciences:** Model networks, optimize resources, and predict outcomes.
- **Everyday Life:** Recommend content on streaming platforms, analyze patterns in social networks, and understand trends in data.

What You Will Learn

By engaging with this course, you will:

1. Solve systems of linear equations using matrices and row operations.
2. Understand the structure and properties of vector spaces.
3. Apply concepts of linear independence, basis, and dimension.
4. Use eigenvalues and eigenvectors to analyze transformations and dynamic systems.
5. Connect theory to applications in physics, engineering, computer science, and beyond.

Linear algebra is not only a set of techniques—it is a way of thinking. It equips you with both the conceptual understanding and practical tools to approach complex problems with clarity and confidence. This book will guide you step by step, blending theory, examples, and applications to help you see both the beauty and utility of linear algebra.

This page titled 0: Why Linear Algebra? is shared under a license and was authored, remixed, and/or curated by LibreTexts.

1: SOLVING SYSTEMS OF LINEAR EQUATIONS

1.1: Systems of Linear Equations

1.2: Matrix Notation and Row Reduction

1.3: Rank and Homogeneous Systems

1.4: Chemistry Reaction-Balancing Chemical Reaction

1.5: Applications to Physics and Resistor Networks

1.6: Chapters Exercises

This page titled 1: Solving 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: Systems of Equations by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

1.1: Systems of Linear Equations

Learning Objectives

- Define linear equations and systems of linear equations.
- Classify systems of linear equations based on the existence and uniqueness of solutions.
- Interpret geometrically two-variable systems as intersecting, parallel, or coinciding lines.
- Extend geometric understanding to three-variable systems, including intersecting, non-intersecting, or coinciding planes.
- Recognize the increasing complexity of systems as the number of variables grows and the need for linear algebra techniques to analyze them.

We will start with a definition of a linear equation. You may have encountered them before, but here is a general definition that we will use.

Definition 1.1.1: Linear

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

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b \tag{1.1.1}$$

where $a_1, a_2, \dots, a_n$ are called **coefficients** and b is called **constant** all are real or complex numbers.

Example 1.1.1

Whic of the following equations is a linear equation

$$3x + 4y = 2z \tag{1.1.2}$$

$$x_1x_2 + 3x_4 = x_3 \tag{1.1.3}$$

Solution

The equation $3x + 4y = 2z$ is linear since it's satisfy the definion bove. However $x_1x_2 + 3x_4 = x_3$ since the two variables x_1 and (x_2 \) are **multiplied together.**

We will usually move the unknowns to the left side of the equation, and move the constants to the right.

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.

Example 1.1.1

Consider the system of linear equations,

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

how many variables are involved in the system

Solution

The system consists of two equations and **three variables**, namely ***x***, ***y***, and ***z***, since each variable appears at least once in the system.

Definition 1.1.4: $\mathbb{R}$

Let n be a positive whole number. We define

$$\mathbb{R}^n = \text{all ordered } n\text{-tuples of real numbers } (x_1, x_2, x_3, \dots, x_n).$$

An n-tuple of real numbers is called a ***point*** of $\mathbb{R}^n$.

In other words, $\mathbb{R}^n$ is just the set of all (ordered) lists of n real numbers. We will draw pictures of $\mathbb{R}^n$ in a moment, but keep in mind that *this is the definition*. For example, $(0, \frac{3}{2}, -\pi)$ and $(1, -2, 3)$ are points of $\mathbb{R}^3$. Also the point $(0, 0, 0, 0, 0)$ is in $\mathbb{R}^5$ and it has a spesific name ***the origin of*** $\mathbb{R}^5$.

Definition 1.1.2: Solution sets

- A ***solution*** of a system of equations is the point $(x_1, x_2, \dots, x_n)$ that make all of the equations true simultaneously.
- The ***solution set*** of a system of equations is the collection of all solutions.
- ***Solving*** the system means finding all of the solutions.

A system of linear equations can have ***one unique solution***, ***infinitely many solutions***, or ***no solution*** at all. Below are the terminologies used to describe the two main categories of solutions for any given system of linear equations.

Definition 1.1.3: Inconsistent and Consistent

A system of equations is called ***inconsistent*** if it has no solutions. It is called ***consistent*** otherwise.

Example 1.1.1

Determine whether the system is consistent or inconsistent:

$$\begin{cases} 2x - y = -1 \\ 2x - y = 3 \end{cases} \tag{1.1.4}$$

\textbf{Solution:}

We will use elimination to solve. First multiply the send equation by negative and then add the two equations

$$\begin{aligned} 2x - y &= -1 \\ -(2x - y &= 3) \\ 0 &= -4 \end{aligned} \tag{1.1.5}$$

\textbf{Conclusion:} Since we arrived at a false statement, there is no point of intersection, and the system is **inconsistent**. Graphically, these lines are parallel. You can plot the two lines to confirm that they never intersect.

Example 1.1.1

Determine whether the following system of equations is consistent or inconsistent. Solve the system using back-substitution.

$$\begin{cases} x + 2y + 3z = 9 \\ y + z = 3 \\ z = 1 \end{cases} \tag{1.1.6}$$

Solution

In this system, we already know the value of z, and the second equation involves only y and z. The fastest way to solve is:

1. Substitute $z = 1$ into the second equation to solve for y:

$$y + 1 = 3 \implies y = 2 \tag{1.1.7}$$

2. Substitute $y = 2$ and $z = 1$ into the first equation to solve for x:

$$x + 2(2) + 3(1) = 9 \implies x + 4 + 3 = 9 \implies x = 2 \tag{1.1.8}$$

Thus, we have the unique point in $\mathbb{R}^3$ as the solution:

$$(x, y, z) = (2, 2, 1) \tag{1.1.9}$$

Thus, the system is **consistent**.

Systems of Equations, Geometry

Outcomes

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

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

When looking for the intersection of lines on a graph, several situations may occur. The following graphs shows the possible cases for two equations (top three graphs) and three equations (bottom two graphs), each involving two variables.

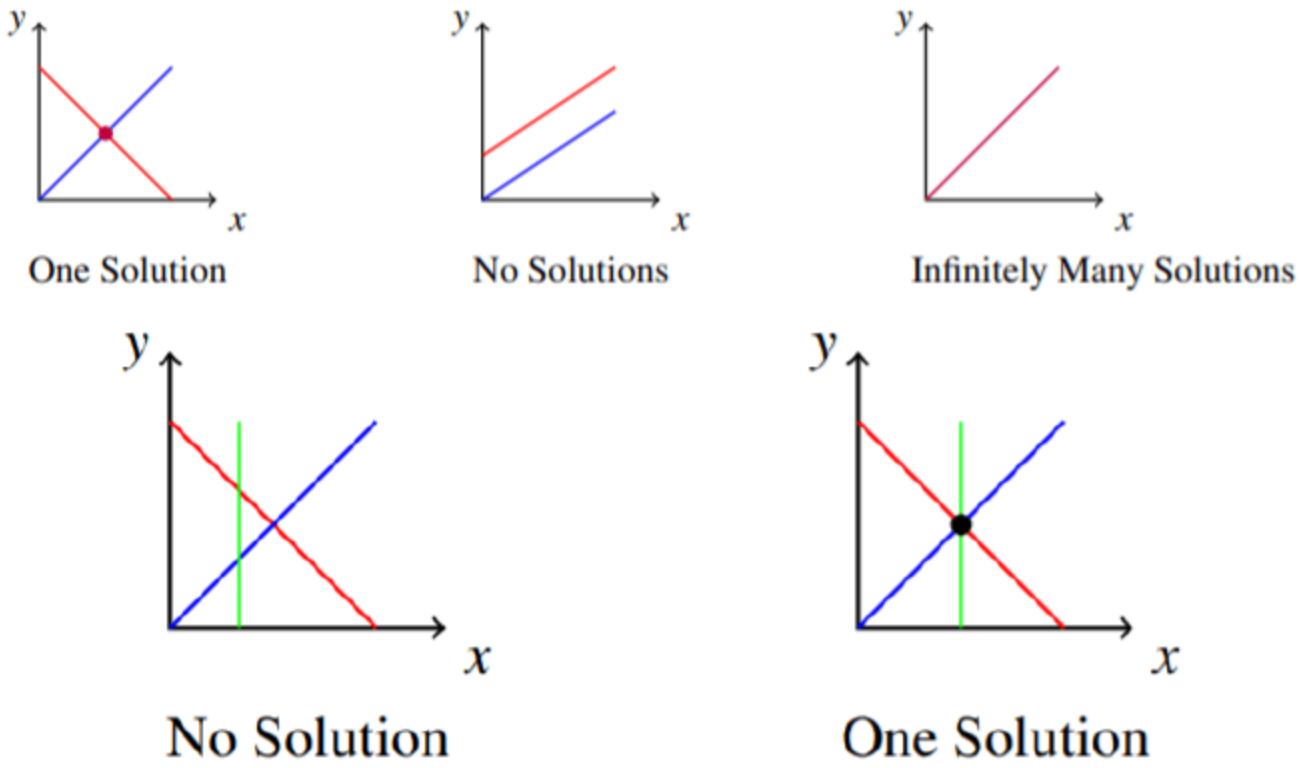


Figure 1.1.1

This illustration shows the possible solutions of a linear system. In the first case, graphs labeled **"one solution"** show lines intersecting at a single point, representing a unique solution to the system. In the second case, graphs labeled **"no solution"** show lines that do not intersect—either because they are parallel or because not all lines meet at the same point—indicating that there is no solution. In the third case, graphs labeled **"infinitely many solutions"** show lines that coincide, meaning they are the same line; here, every point on the line satisfies both equations, resulting in infinitely many solutions.

Example 1.1.1: A Graphical Solution

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

$$x + y = 3$$
$$y - x = 5$$

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

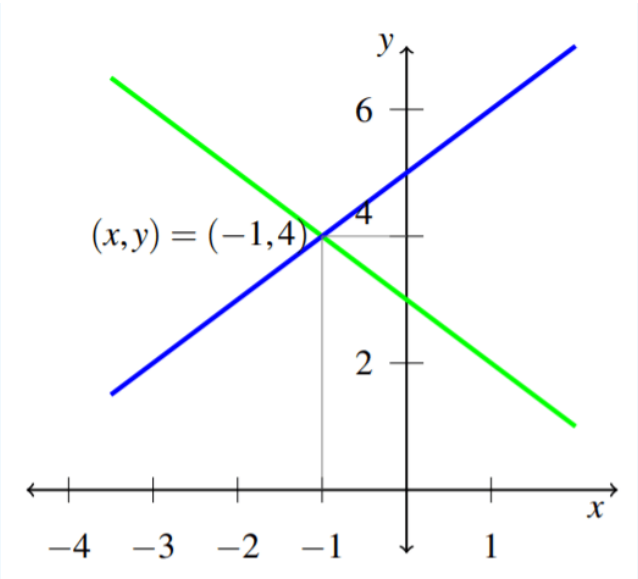


Figure 1.1.2

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 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.

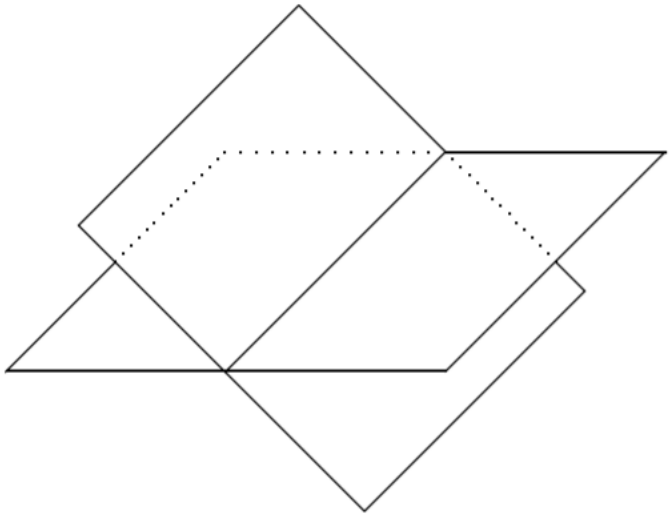


Figure 1.1.4

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.

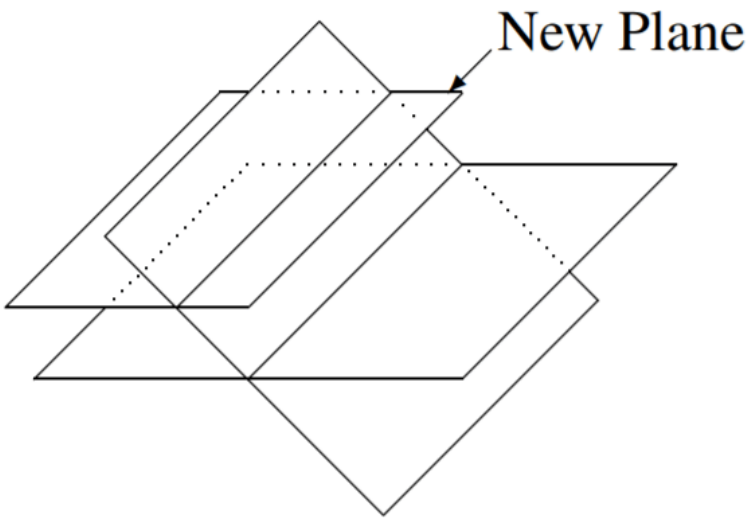


Figure 1.1.5

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.

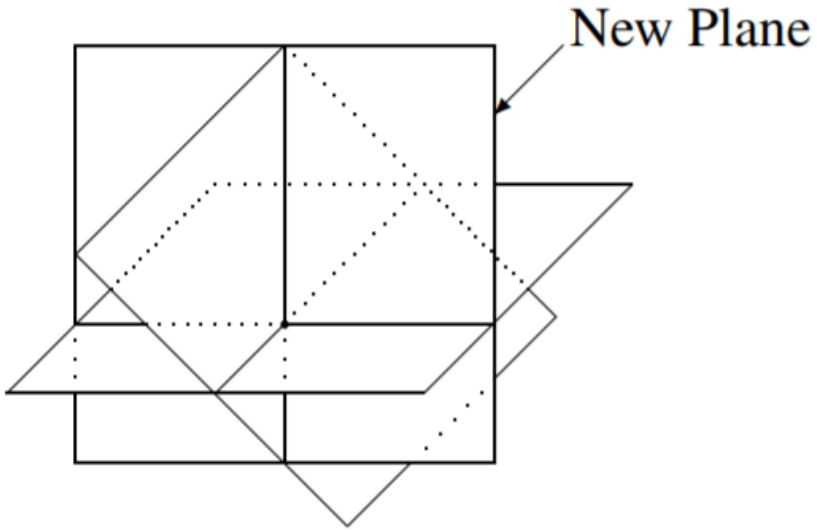


Figure 1.1.6

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.

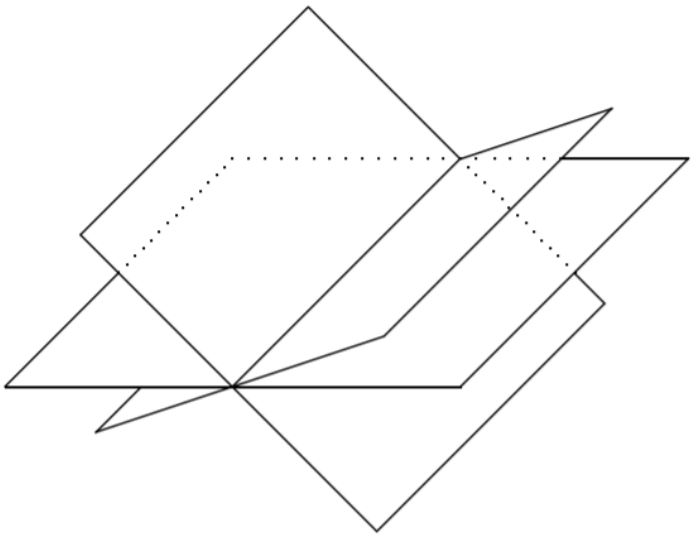


Figure 1.1.7

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.

This page titled 1.1: Systems of Linear Equations is shared under a license and was authored, remixed, and/or curated by LibreTexts.

- 1.1: Systems of Linear Equations by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.
- 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.
- 1.1: Systems of Equations, Geometry by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

1.2: Matrix Notation and Row Reduction

Learning Objectives

- Represent a system of linear equations as an augmented matrix.
- Apply elementary row operations to simplify matrices systematically.
- Transform a matrix into row echelon form (REF) and reduced row echelon form (RREF).
- Identify basic and free variables in a system and understand their implications for solutions.
- Determine whether a system has a unique solution, infinitely many solutions, or no solution.
- Express solutions with parameters when appropriate and solve practical problems using these techniques.

In the previous section, we reviewed definitions and solved systems of equations with two and three variables, much like the methods you encountered in a precalculus course. In this section, we will shift to the linear algebra perspective, where systematic tools—such as matrices and row operations—allow us to solve systems more efficiently and extend these ideas to larger and more complex systems.

Augmented Matrix

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 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 of three variables and three equations

$$\begin{cases} (E1) & 2x+3y-z &= 1 \\ (E2) & 10x-z &= 2 \\ (E3) & 4x-9y+2z &= 5 \end{cases}$$

We encode this system into a matrix by assigning each equation to a corresponding row. Within that row, each variable and the constant gets its own column, and to separate the variables on the left hand side of the equation from the constants on the right hand side, we use a vertical bar, $|$. Note that in $E2$, since y is not present, we record its coefficient as 0. The matrix augmented this system is the following 3×3

$$\begin{array}{c} \\ (E1) \to \\ (E2) \to \\ (E3) \to \end{array} \begin{array}{c} \begin{array}{cccc} x & y & z & b \end{array} \\ \left(\begin{array}{cccc} 2 & 3 & -1 & 1 \\ 10 & 0 & -1 & 2 \\ 4 & -9 & 2 & 5 \end{array}\right) \end{array}$$

Notice that the line separates the coefficient matrix, allowing us to study it separately. We will see that this will be useful in another representation of a linear system (see the note below).

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

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{pmatrix}$$

We will look at more examples of how to write the corresponding augmented matrix for a system of equations, since this is the first step in solving with linear algebra methods. Afterward, we will introduce an algorithm for performing row operations on augmented matrices, which will serve as the foundation for solving systems systematically.

Elementary Row Operations

When solving systems of equations, we often perform certain operations to simplify the equations without changing their solutions. For example, we might swap two equations, multiply an equation by a nonzero constant, or replace one equation with a combination of itself and another. These same operations can be carried out directly on the rows of an augmented matrix. In this way, changes to a system of equations correspond exactly to changes in its augmented matrix, and the solution set remains unchanged. With this motivation, we now introduce the formal definition of **elementary row operations**, which will be our main tool for solving systems systematically.

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.

We first introduce an example to demonstrate **row operations** applied to a system of linear equations and their corresponding **augmented matrix**. Below are some possible row operations that can be used to solve the system. Notice that the **goal is to eliminate variables systematically** to find the solution, and the primary focus here is to **understand the notation** for these operations.

Example 1.2.1

Consider the following system of linear equations:

$$\begin{cases} 3x - y + z &= 3 \\ 2x - 4y + 3z &= 16 \\ x - y + z &= 5 \end{cases} \tag{1.2.1}$$

We start by forming an augmented matrix:

$$\begin{cases} 3x - y + z &= 3 \\ 2x - 4y + 3z &= 16 \\ x - y + z &= 5 \end{cases} \xrightarrow{\text{agumented matrix}} \begin{pmatrix} 3 & -1 & 1 & 3 \\ 2 & -4 & 3 & 16 \\ 1 & -1 & 1 & 5 \end{pmatrix} \tag{1.2.2}$$

Below are possible row operations applied to the system and their corresponding augmented matrices. Notice that we use a single arrow $rightarrow$ to apply a row operation to one row, and a double arrow $leftrightarrow$ when swapping rows.

-Switch two rows.:

$$\begin{cases} (E1) & 3x-y+z &=& 3 \\ (E2) & 2x-4y+3z &=& 16 \\ (E3) & x-y+z &=& 5 \end{cases} \xrightarrow{\text{Switch } E1 \text{ and } E3} \begin{cases} (E1) & x-y+z &=& 5 \\ (E2) & 2x-4y+3z &=& 16 \\ (E3) & 3x-y+z &=& 3 \end{cases}$$

$$\begin{pmatrix} 3 & -1 & 1 & 3 \\ 2 & -4 & 3 & 16 \\ 1 & -1 & 1 & 5 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{pmatrix} 1 & -1 & 1 & 5 \\ 2 & -4 & 3 & 16 \\ 3 & -1 & 1 & 3 \end{pmatrix}$$

-Multiply a row by a nonzero number.

$$\begin{cases} (E1) & 3x_1+x_2+x_4 &=& 6 \\ (E2) & 2x_1+x_2-x_3 &=& 4 \\ (E3) & x_2-3x_3-2x_4 &=& 0 \end{cases} \xrightarrow{\text{Replace } E1 \text{ with } \frac{1}{3}E1} \begin{cases} (E1) & x_1+\frac{1}{3}x_2+\frac{1}{3}x_4 &=& 2 \\ (E2) & 2x_1+x_2-x_3 &=& 4 \\ (E3) & x_2-3x_3-2x_4 &=& 0 \end{cases}$$

$$\begin{pmatrix} 3 & 1 & 0 & 1 & 6 \\ 2 & 1 & -1 & 0 & 4 \\ 0 & 1 & -3 & -2 & 0 \end{pmatrix} \xrightarrow{R_1 \to \frac{1}{3}R_1} \begin{pmatrix} 1 & \frac{1}{3} & 0 & \frac{1}{3} & 2 \\ 2 & 1 & -1 & 0 & 4 \\ 0 & 1 & -3 & -2 & 0 \end{pmatrix}$$

-Replace a row by any multiple of another row added to it.

$$\begin{cases} (E1) & x+\frac{3}{2}y-\frac{1}{2}z &=& \frac{1}{2} \\ (E2) & 10x-z &=& 2 \\ (E3) & 4x-9y+2z &=& 5 \end{cases} \xrightarrow[\text{Replace } E3 \text{ with } -4E1+E3]{\text{Replace } E2 \text{ with } -10E1+E2} \begin{cases} (E1) & x+\frac{3}{2}y-\frac{1}{2}z &=& \frac{1}{2} \\ (E2) & -15y+4z &=& -3 \\ (E3) & -15y+4z &=& 3 \end{cases}$$

$$\begin{pmatrix} 1 & \frac{3}{2} & -\frac{1}{2} & \frac{1}{2} \\ 10 & 0 & -1 & 2 \\ 4 & -9 & 2 & 5 \end{pmatrix} \xrightarrow[R_3 \to -4R_1+R_3]{R_2 \to -10R_1+R_2} \begin{pmatrix} 1 & \frac{3}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & -15 & 4 & -3 \\ 0 & -15 & 4 & 3 \end{pmatrix}$$

To solve a system of linear equations, we can **translate elementary row operations into an elimination method** applied to the augmented matrix of the system. Elementary row operations—swapping rows, multiplying a row by a nonzero scalar, and adding a multiple of one row to another—allow us to systematically **eliminate variables** from equations. By performing these operations strategically, we can transform the augmented matrix into an **upper triangular (staircase) form**, where each successive row contains fewer variables. This triangular form makes the system easier to solve using **back substitution**, starting from the last equation and substituting values upward to find all unknowns. This method not only streamlines computations but also clearly demonstrates the structure and dependencies within the system.
Below is an example showing how to apply row operations to solve a system of linear equations using the elimination method.

Example 1.2.2

Solve the following system of linear equation:

$$\begin{cases} x & + & 2y & + & 3x & = & 6 \\ 2x & - & 3y & + & 2z & = & 14 \\ 3x & + & y & - & z & = & -2 \end{cases} \tag{1.2.3}$$

Solution

We start by forming an augmented matrix:

$$\begin{cases} x & + & 2y & + & 3x & = & 6 \\ 2x & - & 3y & + & 2z & = & 14 \\ 3x & + & y & - & z & = & -2 \end{cases} \xrightarrow{\text{agumented matrix}} \begin{pmatrix} 1 & 2 & 3 & 6 \\ 2 & -3 & 2 & 14 \\ 3 & 1 & -1 & -2 \end{pmatrix}.$$

Eliminating a variable from an equation means producing a zero to the left of the line in an augmented matrix. First we produce zeros in the first column (i.e. we eliminate x) by subtracting multiples of the first row.

$$\begin{pmatrix} 1 & 2 & 3 & 6 \\ 2 & -3 & 2 & 14 \\ 3 & 1 & -1 & -2 \end{pmatrix} \xrightarrow{R_2 \to R_2 - 2R_1} \begin{pmatrix} 1 & 2 & 3 & 6 \\ 0 & -7 & -4 & 2 \\ 3 & 1 & -1 & -2 \end{pmatrix} \xrightarrow{R_3 \to -3R_1 + R_3} \begin{pmatrix} 1 & 2 & 3 & 6 \\ 0 & -7 & -4 & 2 \\ 0 & -5 & -10 & -20 \end{pmatrix} \tag{1.2.4}$$

This was made much easier by the fact that the top-left entry is equal to 1, so we can simply multiply the first row by the number below and subtract. In order to eliminate y in the same way, we would like to produce a 1 in the second column. We could divide the second row by -7, but this would produce fractions; instead, let's divide the third by -5.

$$\begin{pmatrix} 1 & 2 & 3 & 6 \\ 0 & -7 & -4 & 2 \\ 0 & -5 & -10 & -20 \end{pmatrix} \xrightarrow{R_3 \to (1/5)R_3} \begin{pmatrix} 1 & 2 & 3 & 6 \\ 0 & -7 & -4 & 2 \\ 0 & 1 & 2 & 4 \end{pmatrix} \xrightarrow{R_2 \leftrightarrow R_3} \begin{pmatrix} 1 & 2 & 3 & 6 \\ 0 & 1 & 2 & 4 \\ 0 & -7 & -4 & 2 \end{pmatrix} \xrightarrow{R_3 \to R_3 + 7R_2} \begin{pmatrix} 1 & 2 & 3 & 6 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 10 & 30 \end{pmatrix} \xrightarrow{R_3 \to R_3 \div 10} \begin{pmatrix} 1 & 2 & 3 & 6 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 1 & 3 \end{pmatrix} \tag{1.2.5}$$

We swapped the second and third row just to keep things orderly. Now we translate this augmented matrix back into a system of equations:

$$\begin{pmatrix} 1 & 2 & 3 & 6 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 1 & 3 \end{pmatrix} \xrightarrow{\text{becomes}} \begin{cases} x + 2y + 3z = 6 \\ y + 2z = 4 \\ z = 3 \end{cases}$$

Hence $z = 3$; back-substituting as in Example 1.2.1 gives $(x, y, z) = (1, -2, 3)$.

Performing row operations on a matrix **does not change the solution set** of the corresponding system of linear equations. In other words, no matter how we swap rows, scale a row by a nonzero constant, or add a multiple of one row to another, the set of solutions remains exactly the same.

Definition 1.2.1: Row Equivalent

Two matrices are called ***row equivalent*** if one can be obtained from the other by doing some number of row operations.

So the linear equations of row-equivalent matrices have the *same solution set*.

Example 1.2.3: An Inconsistent System

Solve the following system of equations using row operations:

$$\begin{cases} x + y = 2 \\ 3x + 4y = 5 \\ 4x + 5y = 9 \end{cases}$$

Solution

$$\begin{aligned}
\begin{pmatrix} 2 & 10 & -1 \\ 3 & 15 & 2 \end{pmatrix} \quad &\to R_1 \leftarrow R_1 \div 2 \quad \begin{pmatrix} 1 & 5 & -\frac{1}{2} \\ 3 & 15 & 2 \end{pmatrix} \\
&\to R_2 \leftarrow R_2 - 3R_1 \quad \begin{pmatrix} 1 & 5 & -\frac{1}{2} \\ 0 & 0 & \frac{7}{2} \end{pmatrix} \\
&\to R_2 \leftarrow R_2 \times \frac{2}{7} \quad \begin{pmatrix} 1 & 5 & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix} \\
&\to R_1 \leftarrow R_1 + \frac{1}{2}R_2 \quad \begin{pmatrix} 1 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{aligned} \tag{1.2.6}$$

This row-reduced matrix corresponds to the inconsistent system:

$$\begin{cases} x + 5y &= 0 \\ 0 &= 1 \end{cases} \tag{1.2.7}$$

Notice the last column is a pivot column.

This translates back into the system of equations

$$\begin{cases} x + y &= 2 \\ y &= -1 \\ 0 &= 2. \end{cases}$$

Our original system has the same solution set as this system. But this system has no solutions: there are no values of x, y making the third equation true! We conclude that our original equation was inconsistent.

Echelon Forms

In the previous examples, we saw how to translate a system of linear equations into an augmented matrix and apply row operations. The next step is to work directly with matrices, without referencing the original equations. To do this, we need a systematic algorithm for solving an augmented matrix. First, we define what it means for a matrix to be "solved." We say a matrix is in **row echelon form (REF)** when it has a staircase-like structure: each leading entry (the first nonzero number from the left) of a row is to the right of the leading entry in the row above, and all entries below a leading entry are zero. Transforming a matrix into row echelon form using **elementary row operations** provides a clear pathway for solving the system through **back substitution**, allowing us to determine all unknowns systematically and efficiently.

Below is the official definition of row echelon form (REF):

Definition 1.2.2: Row Echelon Form

A matrix is in ***row echelon*** *form* if:

1. All zero rows are at the bottom.
2. The first nonzero entry of a row is to the *right* of the first nonzero entry of the row above.
3. Below the first nonzero entry of a row, all entries are zero.

Here is a picture of a matrix in row echelon form:

$$\begin{pmatrix} \boxed{\star} & \star & \star & \star & \star \\ 0 & \boxed{\star} & \star & \star & \star \\ 0 & 0 & 0 & \boxed{\star} & \star \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{aligned} \star &= \text{any number} \\ \boxed{\star} &= \text{any nonzero number} \end{aligned}$$

A matrix in row-echelon form is generally easy to solve using back-substitution. For example,

$$\begin{pmatrix} 1 & 2 & 3 & 6 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 10 & 30 \end{pmatrix} \xrightarrow{\text{becomes}} \begin{cases} x + 2y + 3z = 6 \\ y + 2z = 4 \\ 10z = 30. \end{cases}$$

We immediately see that $z = 3$, which implies $y = 4 - 2 \cdot 3 = -2$ and $x = 6 - 2(-2) - 3 \cdot 3 = 1$. See Example 1.2.3.

Once a matrix is in **row echelon form (REF)**, we can take the process one step further and transform it into **reduced row echelon form (RREF)**. In RREF, not only does the matrix maintain the staircase-like structure of REF, but each leading entry is also **1** and is the only nonzero entry in its column. By continuing elimination in this way, we can solve the system **directly from the matrix**, without the need for back substitution. The goal is to simplify the matrix so that each row corresponds to a single variable with its value, providing a clear and immediate solution to the system. This approach extends the elimination method fully to a systematic, straightforward procedure for solving any system of linear equations.

Definition 1.2.4: Reduced Row Echelon Form

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

4. Each pivot is equal to 1.
5. Each pivot is the only nonzero entry in its column.

Here is a picture of a matrix in reduced row echelon form:

$$\begin{pmatrix} 1 & 0 & \star & 0 & \star \\ 0 & 1 & \star & 0 & \star \\ 0 & 0 & 0 & 1 & \star \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{aligned} \star &= \text{any number} \\ 1 &= \text{leading term} \end{aligned}$$

A matrix in reduced row echelon form is in some sense completely solved. For example,

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{pmatrix} \xrightarrow{\text{becomes}} \begin{cases} x = 1 \\ y = -2 \\ z = 3. \end{cases}$$

Example 1.2.4

The following matrices are in ***reduced row echelon form***:

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 8 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 17 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The following matrices are in row echelon form but ***not reduced row echelon form***:

$$\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 2 & 7 & 1 & 4 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix} \quad \begin{pmatrix} 1 & 17 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 2 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}.$$

The following matrices are ***not in echelon form***:

$$\begin{pmatrix} 2 & 7 & 1 & 4 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 3 \end{pmatrix} \quad \begin{pmatrix} 0 & 17 & 0 \\ 0 & 2 & 1 \end{pmatrix} \quad \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Example 1.2.5

Reduce the following augmented matrix to solve for the corresponding system of linear equation:

$$\begin{pmatrix} 0 & -7 & -4 & 2 \\ 2 & 4 & 6 & 12 \\ 3 & 1 & -1 & -2 \end{pmatrix}.$$

Solution

The reduced row echelon form of the matrix is

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{pmatrix} \xrightarrow{\text{translates to}} \begin{cases} x & & & = & 1 \\ & y & & = & -2 \\ & & z & = & 3. \end{cases}$$

The reduced row echelon form of the matrix tells us that the only solution is $(x, y, z) = (1, -2, 3)$.

Figure 1.2.2: Animated slideshow of the row reduction in this example.

Here is the row reduction algorithm, summarized in pictures.

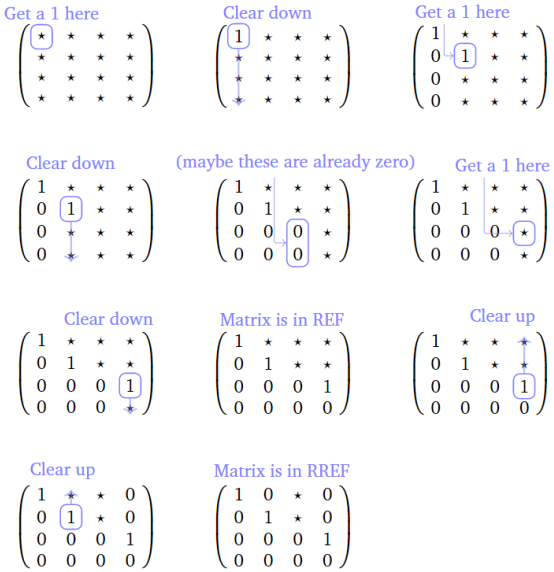


Figure 1.2.3

After working through examples of transforming a matrix into reduced row echelon form, we may begin to wonder whether the final result depends on the sequence of row operations we choose. Different paths can certainly be taken, but the remarkable fact is that all lead to the same destination. This observation is formalized in the following theorem, which guarantees the **uniqueness of the reduced echelon form** of a matrix.

Theorem 1.2.2: Uniqueness of the Reduced Row-Echelon Form

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

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

Pivot Position, Basic and Free variables

Once we obtain a matrix in reduced row echelon form (RREF), it becomes much easier to analyze the structure of the solutions. Recognizing the columns that contain leading entries is important because they correspond to the variables determined directly by the system. We call these positions **pivot positions**, and they provide a systematic way to describe the solution set of the system.
Now, looking back at the example above, notice that our goal is to create **pivot columns** and identify the corresponding **pivot positions**.

Definition 1.2.5

A ***pivot position*** of a matrix is an entry that is a pivot of a row echelon form of that matrix.

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

Example 1.2.6: Pivot Positions

Find the pivot positions and pivot columns of this matrix

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

Solution

We saw in Example 1.2.5 that a row echelon form of the matrix is

$$\begin{pmatrix} 1 & 2 & 3 & 6 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 10 & 30 \end{pmatrix}.$$

The pivot positions of A are the entries that become pivots in a row echelon form; they are marked in red below:

$$\begin{pmatrix} 0 & -7 & -4 & 2 \\ 2 & 4 & 6 & 12 \\ 3 & 1 & -1 & -2 \end{pmatrix}.$$

The first, second, and third columns are pivot columns.

Once we have identified the pivot positions in the reduced row echelon form of a matrix, we can use them to describe the solutions of a system of linear equations. Pivot positions correspond to the variables that are directly determined by the system, while the remaining variables—those without pivots—are called ***free variables***. The relationship between pivot and free variables provides a systematic way to express the solution set, whether it consists of a single unique solution or infinitely many. However, not all systems have solutions.

In other words, the row reduced matrix of an inconsistent system looks like this:

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

In the next example, we will see how an inconsistent system can be recognized from its reduced form and the arrangement of pivot columns.

Example 1.2.7: An Inconsistent System

Solve the linear system

$$\begin{cases} 2x + 10y = -1 \\ 3x + 15y = 2 \end{cases}$$

using row reduction.

Solution

$$\begin{aligned} \begin{pmatrix} 2 & 10 & -1 \\ 3 & 15 & 2 \end{pmatrix} &\xrightarrow{R_1 \leftarrow R_1 \div 2} \begin{pmatrix} 1 & 5 & -\frac{1}{2} \\ 3 & 15 & 2 \end{pmatrix} \\ &\xrightarrow{R_2 \leftarrow R_2 - 3R_1} \begin{pmatrix} 1 & 5 & -\frac{1}{2} \\ 0 & 0 & \frac{7}{2} \end{pmatrix} \\ &\xrightarrow{R_2 \leftarrow R_2 \times \frac{2}{7}} \begin{pmatrix} 1 & 5 & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix} \\ &\xrightarrow{R_1 \leftarrow R_1 + \frac{1}{2}R_2} \begin{pmatrix} 1 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned} \tag{1.2.8}$$

This row-reduced matrix corresponds to the inconsistent system:

$$\begin{cases} x + 5y = 0 \\ 0 = 1 \end{cases} \tag{1.2.9}$$

In the above example, we saw how recognizing the reduced row echelon form of an inconsistent system leads us to the following theorem:

Theorem 1.2.2: The Row Echelon Form of an Inconsistent System

An augmented matrix corresponds to an inconsistent system of equations if and only if the last column (i.e., the augmented column) is a *pivot column*.

Not having a pivot column in every column—or equivalently, not having enough independent equations to assign a pivot to each variable—means that some variables will not be determined directly by the system. These undetermined variables are called free variables, as mentioned previously. The variables corresponding to pivot columns, which are determined directly by the system, are called ***basic variables***. The presence of free variables indicates that the system does not have a unique solution; instead, it has infinitely many solutions. By expressing the basic (pivot) variables in terms of these free variables, we can systematically describe all possible solutions and understand the structure of the solution set.

Here are some example

Example 1.2.8: A System with Many Solutions

Solve the linear system

$$\begin{cases} 2x + y + 12z = 1 \\ x + 2y + 9z = -1 \end{cases}$$

using row reduction, while identifying pivot positions, basic variables, and free variables.

\textbf{Solution:}

$$\begin{aligned}
&\begin{pmatrix} 2 & 1 & 12 & 1 \\ 1 & 2 & 9 & -1 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{pmatrix} 1 & 2 & 9 & -1 \\ 2 & 1 & 12 & 1 \end{pmatrix} && \text{(Step 1: first pivot)} \\
&\xrightarrow{R_2 \to R_2 - 2R_1} \begin{pmatrix} 1 & 2 & 9 & -1 \\ 0 & -3 & -6 & 3 \end{pmatrix} && \text{(Step 2: eliminate below pivot)} \\
&\xrightarrow{R_2 \to (-1/3)R_2} \begin{pmatrix} 1 & 2 & 9 & -1 \\ 0 & 1 & 2 & -1 \end{pmatrix} && \text{(Step 3: scale second pivot)} \\
&\xrightarrow{R_1 \to R_1 - 2R_2} \begin{pmatrix} 1 & 0 & 5 & 1 \\ 0 & 1 & 2 & -1 \end{pmatrix} && \text{(Step 4: eliminate above pivot)}
\end{aligned} \tag{1.2.10}$$

This row-reduced matrix corresponds to the linear system

$$\begin{cases} x + 5z = 1 \\ y + 2z = -1 \end{cases}$$

Analysis of solutions:

- **Pivot positions:** The first entry in row 1 (column 1) and the first entry in row 2 (column 2) are pivot positions.
- **Basic variables:** The variables corresponding to pivot columns, x and y, are basic variables.
- **Free variable:** Column 3 does not contain a pivot, so z is a free variable.

- **Infinitely many solutions:** Because there is a free variable, the system has infinitely many solutions. Expressing the basic variables in terms of the free variable z:
  $x = 1 - 5z, y = -1 - 2z, z = z$ (free).

To describe all solutions of a system with free variables more conveniently, we can introduce a parameter (a parameter represnt any real number) , often denoted by t, to represent the free variable. This allows us to write the basic variables explicitly in terms of tt, giving a single, unified expression for every solution in the solution set. By parametrizing the free variable, we can clearly see how the free variable affects the values of the basic variables and conveniently describe the infinitely many solutions in a compact form.

Continuing with the previous example, we can let the free variable z be represented by the parameter t. Then the solution set can be written as:

$$x = 1 - 5t, \quad y = -1 - 2t, \quad z = t, \quad t^{\text{int}}\mathbb{R}. \tag{1.2.11}$$

This parametrization captures all infinitely many solutions of the system, showing how each choice of t gives a corresponding set of values for the basic variables x and y.

Example 1.2.1: 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

The row-echelon form of the augmented matrix of this system is given by

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

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,

Example 1.2.7: An Infinite Set of Solutions

Give the complete solution to the system of equations

$$\begin{aligned} 3x - y - 5z &= 9 \\ y - 10z &= 0 \\ -2x + y &= -6 \end{aligned} \tag{1.2.12}$$

Solution

The augmented matrix of this system is

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

In order to find the solution to this system, we will carry the augmented matrix to reduced row-echelon form. 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

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

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.

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

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.

$$\begin{bmatrix} 1 & -\frac{1}{3} & -\frac{5}{3} & 3 \\ 0 & 1 & -10 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

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.

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

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

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

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

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

We wish to carry this matrix to row-echelon form. Here, we will outline the row operations used.

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

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

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

$$\begin{bmatrix} 1 & 2 & -1 & 1 & 3 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{1.2.13}$$

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

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

It is customary to write this solution in the form

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} -1 + s - t \\ 2 \\ s \\ t \end{bmatrix} \tag{1.2.14}$$

Note on RREF Algorithm

- Start with the augmented matrix.
- From left to right, locate the first nonzero column. This is the **pivot column**, and its top entry is the **pivot position**. If needed, swap rows so a nonzero entry is in the pivot position.
- Use row operations to make all entries **below** the pivot equal to zero.
- Repeat steps 2–3 with the remaining submatrix, moving downward and to the right, until no rows remain. At this point the matrix is in **row-echelon form**.
- Scale each row so that the leading entry (pivot) in each nonzero row is **1**.
- Finally, working from right to left, use row operations to create zeros **above** each pivot. The result is the **reduced row-echelon form (RREF)**.

Exercises

In Exercises 1 - 6, state whether the given matrix is in reduced row echelon form, row echelon form only or in neither of those forms.

1. $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 3 \end{bmatrix}$
2. $\begin{bmatrix} 3 & -1 & 1 & 3 \\ 2 & -4 & 3 & 16 \\ 1 & -1 & 1 & 5 \end{bmatrix}$
3. $\begin{bmatrix} 1 & 1 & 4 & 3 \\ 0 & 1 & 3 & 6 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
4. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

5. $\left[\begin{array}{ccccc} 1 & 0 & 4 & 3 & 0 \\ 0 & 1 & 3 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

6. $\left[\begin{array}{cccc} 1 & 1 & 4 & 3 \\ 0 & 1 & 3 & 6 \end{array}\right]$

In Exercises 7 - 12, the following matrices are in reduced row echelon form. Determine the solution of the corresponding system of linear equations or state that the system is inconsistent.

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

8. $\left[\begin{array}{cccc} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 20 \\ 0 & 0 & 1 & 19 \end{array}\right]$

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

10. $\left[\begin{array}{ccccc} 1 & 0 & 0 & 3 & 0 \\ 0 & 1 & 2 & 6 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]$

11. $\left[\begin{array}{ccccc} 1 & 0 & -8 & 1 & 7 \\ 0 & 1 & 4 & -3 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

12. $\left[\begin{array}{cccc} 1 & 0 & 9 & -3 \\ 0 & 1 & -4 & 20 \\ 0 & 0 & 0 & 0 \end{array}\right]$

In Exercises 13 - 26, solve the following systems of linear equations using the techniques discussed in this section. Compare and contrast these techniques with those you used to solve the systems in the Exercises in Section 8.1.

13. $\left\{\begin{array}{rcl} -5x+y &=& 17 \\ x+y &=& 5 \end{array}\right.$

14. $\left\{\begin{array}{rcl} x+y+z &=& 3 \\ 2x-y+z &=& 0 \\ -3x+5y+7z &=& 7 \end{array}\right.$

15. $\left\{\begin{array}{rcl} 4x-y+z &=& 5 \\ 2y+6z &=& 30 \\ x+z &=& 5 \end{array}\right.$

16. $\left\{\begin{array}{rcl} x-2y+3z &=& 7 \\ -3x+y+2z &=& -5 \\ 2x+2y+z &=& 3 \end{array}\right.$

17. $\left\{\begin{array}{rcl} 3x-2y+z &=& -5 \\ x+3y-z &=& 12 \\ x+y+2z &=& 0 \end{array}\right.$

18. $\left\{\begin{array}{rcl} 2x-y+z &=& -1 \\ 4x+3y+5z &=& 1 \\ 5y+3z &=& 4 \end{array}\right.$

19. $\left\{\begin{array}{rcl} x-y+z &=& -4 \\ -3x+2y+4z &=& -5 \\ x-5y+2z &=& -18 \end{array}\right.$

20. $\left\{\begin{array}{rcl} 2x-4y+z &=& -7 \\ x-2y+2z &=& -2 \\ -x+4y-2z &=& 3 \end{array}\right.$

21. $\begin{cases} 2x - y + z &= 1 \\ 2x + 2y - z &= 1 \\ 3x + 6y + 4z &= 9 \end{cases}$

22. $\begin{cases} x - 3y - 4z &= 3 \\ 3x + 4y - z &= 13 \\ 2x - 19y - 19z &= 2 \end{cases}$

23. $\begin{cases} x + y + z &= 4 \\ 2x - 4y - z &= -1 \\ x - y &= 2 \end{cases}$

24. $\begin{cases} x - y + z &= 8 \\ 3x + 3y - 9z &= -6 \\ 7x - 2y + 5z &= 39 \end{cases}$

25. $\begin{cases} 2x - 3y + z &= -1 \\ 4x - 4y + 4z &= -13 \\ 6x - 5y + 7z &= -25 \end{cases}$

26. $\begin{cases} x_1 - x_3 &= -2 \\ 2x_2 - x_4 &= 0 \\ x_1 - 2x_2 + x_3 &= 0 \\ -x_3 + x_4 &= 1 \end{cases}$

27. It's time for another meal at our local buffet. This time, 22 diners (5 of whom were children) feasted for $\$162.25$, before taxes. If the kids buffet is $\$4.50$, the basic buffet is $\$7.50$, and the deluxe buffet (with crab legs) is $\$9.25$, find out how many diners chose the deluxe buffet.
28. Carl wants to make a party mix consisting of almonds (which cost $\$7$ per pound), cashews (which cost $\$5$ per pound), and peanuts (which cost $\$2$ per pound.) If he wants to make a 10 pound mix with a budget of $\$35$, what are the possible combinations almonds, cashews, and peanuts? (You may find it helpful to review Example 8.1.3 in Section 8.1.)
29. Find the quadratic function passing through the points $(-2, 1)$, $(1, 4)$, $(3, -2)$
30. At 9 PM, the temperature was 60°F; at midnight, the temperature was 50°F; and at 6 AM, the temperature was 70°F . Use the technique in Example 8.2.3 to fit a quadratic function to these data with the temperature, T, measured in degrees Fahrenheit, as the dependent variable, and the number of hours after 9 PM, t, measured in hours, as the independent variable. What was the coldest temperature of the night? When did it occur?
31. The price for admission into the Stitz-Zeager Sasquatch Museum and Research Station is $15 for adults and $8 for kids 13 years old and younger. When the Zahlenreich family visits the museum their bill is $38 and when the Nullsatz family visits their bill is $39. One day both families went together and took an adult babysitter along to watch the kids and the total admission charge was $92. Later that summer, the adults from both families went without the kids and the bill was $45. Is that enough information to determine how many adults and children are in each family? If not, state whether the resulting system is inconsistent or consistent dependent. In the latter case, give at least two plausible solutions.
32. Use the technique in Example 8.2.3 to find the line between the points $(-3, 4)$ and $(6, 1)$. How does your answer compare to the slope-intercept form of the line in Equation 23?
33. With the help of your classmates, find at least two different row echelon forms for the matrix

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 12 & 8 \end{bmatrix}$$

Answers

1. Reduced row echelon form
2. Neither
3. Row echelon form only
4. Reduced row echelon form
5. Reduced row echelon form
6. Row echelon form only
7. $(-2, 7)$
8. $(-3, 20, 19)$
9. $(-3t + 4, -6t - 6, 2, t)$
   for all real numbers t

10. Inconsistent

11. $(8s - t + 7, -4s + 3t + 2, s, t)$
for all real numbers s and t

12. $(-9t - 3, 4t + 20, t)$
for all real numbers t

13. $(-2, 7)$
14. $(1, 2, 0)$

15. $(-t + 5, -3t + 15, t)$
for all real numbers t

16. $(2, -1, 1)$
17. $(1, 3, -2)$
18. Inconsistent
19. $(1, 3, -2)$
20. $(-3, \frac{1}{2}, 1)$
21. $(\frac{1}{3}, \frac{2}{3}, 1)$

22. $(\frac{19}{13}t + \frac{51}{13}, -\frac{11}{13}t + \frac{4}{13}, t)$
for all real numbers t

23. Inconsistent
24. $(4, -3, 1)$

25. $(-2t - \frac{35}{4}, -t - \frac{11}{2}, t)$
for all real numbers t

26. $(1, 2, 3, 4)$
27. This time, 7 diners chose the deluxe buffet.
28. If t represents the amount (in pounds) of peanuts, then we need $1.5t - 7.5$ pounds of almonds and $17.5 - 2.5t$ pounds of cashews. Since we can't have a negative amount of nuts, $5 \leq t \leq 7$.
29. $f(x) = -\frac{4}{5}x^2 + \frac{1}{5}x + \frac{23}{5}$
30. $T(t) = \frac{20}{27}t^2 - \frac{50}{9}t + 60$. Lowest temperature of the evening $\frac{595}{12} \approx 49.58^\circ$F at 12:45 AM.
31. Let x_1 and x_2 be the numbers of adults and children, respectively, in the Zahlenreich family and let x_3 and x_4 be the numbers of adults and children, respectively, in the Nullsatz family. The system of equations determined by the given information is

$$\begin{cases} 15x_1 + 8x_2 &= 38 \\ 15x_3 + 8x_4 &= 39 \\ 15x_1 + 8x_2 + 15x_3 + 8x_4 &= 77 \\ 15x_1 + 15x_3 &= 45 \end{cases}$$

We subtracted the cost of the babysitter in E3 so the constant is 77, not 92. This system is consistent dependent and its solution is $(\frac{8}{15}t + \frac{2}{5}, -t + 4, -\frac{8}{15}t + \frac{13}{5}, t)$. Our variables represent numbers of adults and children so they must be whole numbers. Running through the values $t = 0, 1, 2, 3, 4$ yields only one solution where all four variables are whole numbers; $t = 3$ gives us $(2, 1, 1, 3)$. Thus there are 2 adults and 1 child in the Zahlenreichs and 1 adult and 3 kids in the Nullsatzs.

This page titled 1.2: Matrix Notation and Row Reduction is shared under a license and was authored, remixed, and/or curated by LibreTexts.

- 1.3: Gaussian Elimination by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.
- 8.2: Systems of Linear Equations- Augmented Matrices by Carl Stitz and Jeff Zeager is licensed CC BY-NC-SA 3.0. Original source: https://www.stitz-zeager.com/latex-source-code.html.
- 1.2: Row Reduction by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.
- 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.

1.3: Rank and Homogeneous Systems

Learning Objectives

- Recognize a homogeneous system of linear equations and identify the trivial solution.
- Determine conditions under which a homogeneous system has nontrivial solutions.
- Understand the concept of the rank of a matrix and how it relates to pivot columns.
- Use the rank of a matrix to analyze the number and type of solutions in a homogeneous system.
- Relate the number of equations and variables to the existence of nontrivial solutions.

Up to this point, we have studied general systems of linear equations and seen how pivots determine whether a system has a unique solution, infinitely many solutions, or no solution. In this section, we will focus on a special type of linear system and explore a crucial characteristic of matrices: how their rank, closely related to the positions of pivots, governs the solutions of the system.

Homogeneous Systems

There is a special type of system which requires additional study. This type of system is called a homogeneous system of equations.

Definition: Homogeneous System of Linear 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.

It is easy to see that this system is consistent, with the origin as a solution. This solution has a special name, which we introduce in the following definition.

Definition 1.3.1: Trivial Solution

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.

Consider the following example.

Example 1.3.1: 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

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

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

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

The corresponding system of equations is

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

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

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

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

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

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

can be written as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

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 1.3.1 has the basic solution $X_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

We explore this further in the following example.

Example 1.3.2: Basic Solutions of a Homogeneous System

Consider the following homogeneous system of equations.

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

Find the basic solutions to this system.

Solution

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

$$\begin{bmatrix} 1 & 4 & 3 & 0 \\ 3 & 12 & 9 & 0 \end{bmatrix} \to \cdots \to \begin{bmatrix} 1 & 4 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

When written in equations, this system is given by

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

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

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

which can be written as

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

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

$$X_1 = \begin{bmatrix} -4 \\ 1 \\ 0 \end{bmatrix}, X_2 = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$

Rank

Another way to gain insight into the solutions of a homogeneous system is to examine the number of pivots in its associated coefficient matrix. This leads us to the concept of **rank**, which we now define as a measure of the number of linearly independent rows (or columns) in a matrix.

Definition 1.3.3: Rank of a Matrix

Let A be a matrix and consider any reduced row-echelon form of A. Then, the number r of pivot columns of A does not depend on the reduced row-echelon form you choose, and is called the **rank** of A. We denote it by Rank(A).

Similarly, we could count the number of leading entries in the row echelon form of A to determine the rank of A.

Example 1.3.3: 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 1.3.1: 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.

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

Example 1.3.4: 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.

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

Theorem 1.3.2: 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. Then there are infinitely many solutions. There are less pivot positions (and hence less leading entries) than columns, meaning that not every column is a pivot column. The columns which are not pivot columns correspond to parameters. In fact, in this case we have $n - r$ parameters.

Exercises

Exercise 1

Find the rank of the following matrix.

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

Answer available at the end of this section.

Exercise 2

Find the rank of the following matrix.

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

Exercise 3

Find the rank of the following matrix.

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

Answer available at the end of this section.

Exercise 4

Find the rank of the following matrix.

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

Exercise 5

Find the rank of the following matrix.

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

Answer available at the end of this section.

Exercise 6

Find the rank of the following matrix.

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

Exercise 7

Find the rank of the following matrix.

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

Answer available at the end of this section.

Exercise 8

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 9

Find the rank of the following matrix.

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

Answer available at the end of this section.

Exercise 10

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 11

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 12

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

Answer available at the end of this section.

Exercise 13

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 available at the end of this section.

Exercise 14

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 available at the end of this section.

Answers

Exercise 1

Rank = 2

Exercise 3

Rank = 3

Exercise 5

Rank = 3

Exercise 7

Rank = 3

Exercise 9

Rank = 2

Exercise 12

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 13

a. This says that both A and $A|B$ have exactly one row of zeros at the bottom. This means a solution exists. Since A has 5 columns but only 4 pivots, the system will have at least one free variable, leading to an infinite solution set.
b. This surely can't happen. The rank cannot get smaller after augmenting another column since we can't *lose* a pivot position in doing this.
c. This is impossible, since A with only 2 columns cannot possibly have 4 pivots.
d. Since $A|B$ has one more pivot than A, in the reduced row echelon form, $A|B$'s last row must be $\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$. 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. In the reduced row echoelon form of both A and $A|B$, the last two rows are rows of 0s.

Exercise 14

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

This page titled 1.3: Rank and Homogeneous Systems is shared under a license and was authored, remixed, and/or curated by LibreTexts.

- 1.5: Rank and Homogeneous Systems 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.
- 1.9: Exercises by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

1.4: Chemistry Reaction-Balancing Chemical Reaction

Learning Objectives

- Create a homogenous system of linear equations to balance a chemical equations based on the Mass Balance Theory.
- Understand that the system will have infinitely many solutions and not only the trivial solution.
- Find the solution such that it is the least positive integer that balance the chemical reaction.

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

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

The reduced row-echelon form of this matrix is

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

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

$$x = \frac{1}{2}t$$
$$y = t$$
$$z = \frac{1}{2}t$$
$$w = t$$

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 1.4.1: 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

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

and the reduced row-echelon form is

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

The solution is given by

$$x - w = 0$$
$$y - \frac{1}{3}w = 0$$
$$z - \frac{1}{3}w = 0$$

which can be written as

$$x = t$$
$$y = \frac{1}{3}t$$
$$z = \frac{1}{3}t$$
$$w = t$$

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.

Exercises

Exercise 1

Balance the following chemical reaction.

$C_3H_8 + O_2 \rightarrow CO_2 + H_2O$

Answer available at the end of this section.

Exercise 2

Balance the following chemical reaction.

$KNO_3 + H_2CO_3 \rightarrow K_2CO_3 + HNO_3$

Exercise 3

Balance the following chemical reaction.

$C_2H_5OH + O_2 \rightarrow CO_2 + H_2O$

Answer available at the end of this section.

Exercise 3

Balance the following chemical reaction.

$AgI + Na_2S \rightarrow Ag_2S + NaI$

Exercise 4

Balance the following chemical reaction.

$P_4O_{10} + H_2O \rightarrow H_3PO_4$

Answer available at the end of this section.

Answers

Exercise 1

$C_3H_8 + 5O_2 \rightarrow 5CO_2 + 5H_2O$

Exercise 3

$C_2H_5OH + 3O_2 \rightarrow 2CO_2 + 3H_2O$

Exercise 4

$P_4O_{10} + 6H_2O \rightarrow 4H_3PO_4$

This page titled 1.4: Chemistry Reaction-Balancing Chemical Reaction is shared under a license and was authored, remixed, and/or curated by LibreTexts.

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

1.5: Applications to Physics and Resistor Networks

Learning Objectives

- Understand the pasts of an electric circuit.
- Write a system of linear equations to find the value and the direction current in a electric circuits use Kirchhoff's law.
- Solve that system by finding the **RREF** of a augmented matrix to the system.
- Understand that this system should be consistent with a unique solution and interpret the solution.

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.

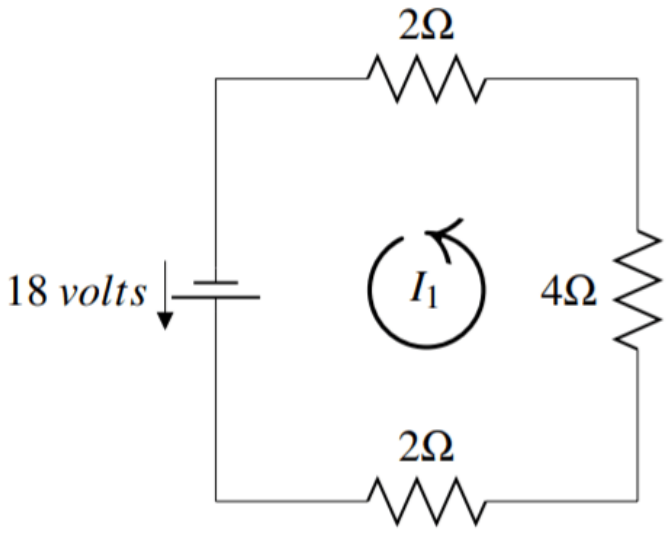


Figure $1.5.1$ A circuit with a voltage source of 18 volts that forms a closed rectangular path with three resistors and an arrow indicating current flow. The resistors have resistance 2Ω, 4Ω, and 2Ω, as we move clockwise from the voltage source. The current flow, I_1, is counterclockwise.

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 Figure $1.5.1$, 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 1.5.1: 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 1.5.1: Solving for Current

Applying Kirchhoff's Law to the circuit in Figure 1.5.1, determine the value for I_1.

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

$$\begin{aligned} 2I_1 + 4I_1 + 2I_1 &= 18 \\ 8I_1 &= 18 \\ I_1 &= \frac{9}{4}A \end{aligned} \tag{1.5.1}$$

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

Example 1.5.2 Solving for Current

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

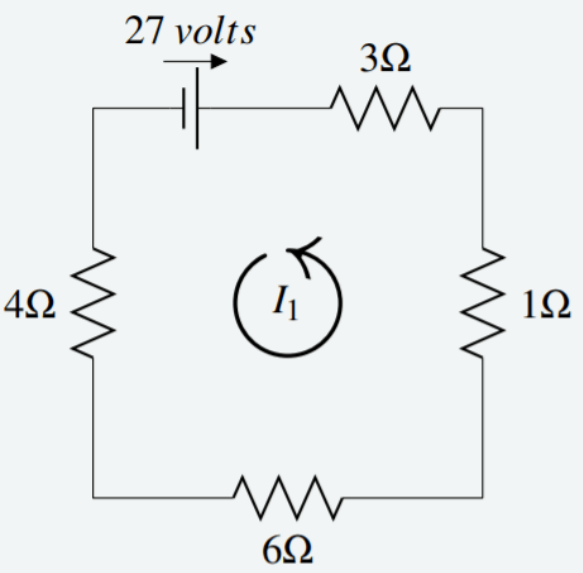


Figure 1.5.2 A circuit with a voltage source of 27 volts that forms a closed rectangular path with four resistors and an arrow indicating current flow. The resistors have resistance 3Ω, 1Ω, 6Ω, and 4Ω, as we move clockwise from the voltage source. The current flow, I_1, is counterclockwise.

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

$$\begin{aligned} 4I_1 + 6I_1 + 1I_1 + 3I_1 &= -27 \\ 14I_1 &= -27 \\ I_1 &= -\frac{27}{14} A \end{aligned} \tag{1.5.2}$$

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 1.5.3: 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.

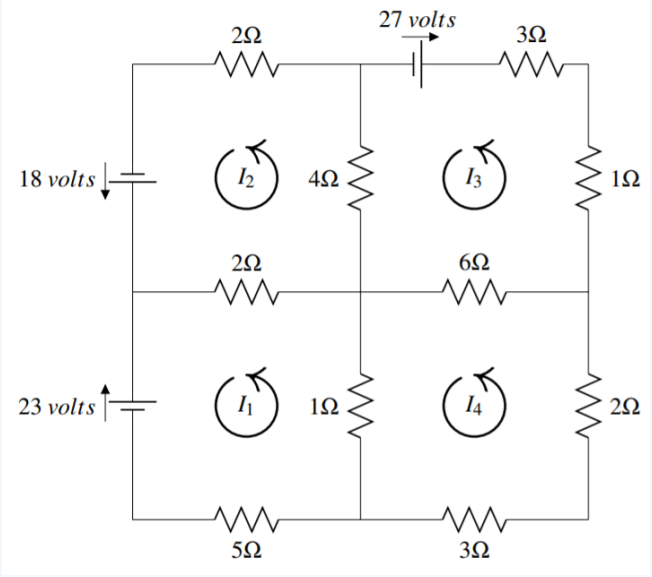


Figure 1.5.3 A circuit with multiple interconnected components, including resistors, voltage sources, and current loops. The circuit is arranged in a 2x2 grid, dividing the circuit into 4 squares: Square 1 on the bottom left, Square 2 on the top left, Square 3 on the top right and Square 4 on the bottom right. Square 1 has counterclockwise current flow I_1, with a 23V voltage source with arrow pointing up, and resistors 5Ω along the bottom, 1Ω on the right shared with Square 4, and 2Ω above shared with Square 2. Square 2 has counterclockwise current flow I_2, with an 18V voltage source with arrow pointing down, and resistors 2Ω on the bottom shared with Square 1, 4Ω on the right, shared iwth Square 3, and 2Ω above. Square 3 has counterclockwise current flow I_3, with resistors 4Ω on the left shared with Square 2, 6Ω on the bottom shared with Square 4, 1Ω on the right, 3Ω above, and a 27V voltage source with arrow pointing right. Square 4 has counterclockwise current flow I_4, with resistors 1Ω on the left shared with Square 1, 3Ω below, 2Ω on the right and 6Ω above shared with Square 3.

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

$$\begin{aligned} 2I_2 - 2I_1 + 4I_2 - 4I_3 + 2I_2 &= 18 \\ 4I_3 - 4I_2 + 6I_3 - 6I_4 + I_3 + I_3 &= -27 \\ 2I_4 + 3I_4 + 6I_4 - 6I_3 + I_4 - I_1 &= 0 \\ 5I_1 + I_1 - I_4 + 2I_1 - 2I_2 &= -23 \end{aligned} \tag{1.5.3}$$

Simplifying and rearranging with variables in order, we have:

$$\begin{aligned} -2I_1 + 8I_2 - 4I_3 &= 18 \\ -4I_2 + 14I_3 - 6I_4 &= -27 \\ -I_1 - 6I_3 + 12I_4 &= 0 \\ 8I_1 - 2I_2 - I_4 &= -23 \end{aligned} \tag{1.5.4}$$

The augmented matrix is

$$\begin{bmatrix} -2 & 8 & -4 & 0 & 18 \\ 0 & -4 & 14 & -6 & -27 \\ -1 & 0 & -6 & 12 & 0 \\ 8 & -2 & 0 & -1 & -23 \end{bmatrix}$$

The solution to this matrix is

$$\begin{aligned} I_1 &= -3A \\ I_2 &= \frac{1}{4}A \\ I_3 &= -\frac{5}{2}A \\ I_4 &= -\frac{3}{2}A \end{aligned} \tag{1.5.5}$$

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

Dimensionless Variables

This section shows how solving systems of equations can be used to determine appropriate dimensionless variables. It is only an introduction to this topic and considers a specific example of a simple airplane wing shown below. We assume for simplicity that it is a flat plane at an angle to the wind which is blowing against it with speed V as shown in Figure $1.5.4$.

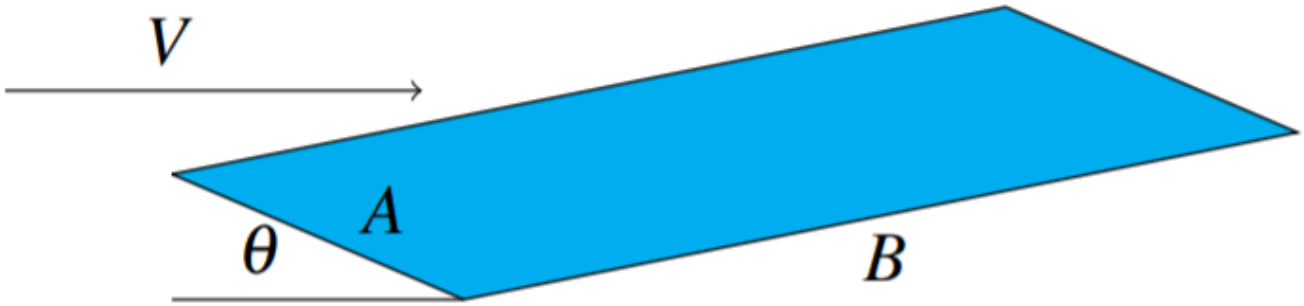


Figure $1.5.4$ The airplane, represented as a slanted rectangle with sides A and B, is slanted to the right. The horizontal arrow on the left, labeled V and pointing towards the airplane, indicates wind speed. The slanted side of the rectangle, labeled A, makes an angle θ with the horizontal, whereas the other side of the rectangle, labeled B, is horizontal.

The angle θ is called the angle of incidence, B is the span of the wing and A is called the chord. Denote by l the lift. Then this should depend on various quantities like θ, V, B, A and so forth. Here is a table which indicates various quantities on which it is reasonable to expect l to depend.

Table $1.5.1$ *Quantities with corresponding symbol and units*

Variable	Symbol	Units
chord	A	m
span	B	m

Variable	Symbol	Units
angle incidence	θ	$m^0kg^0sec^0$
speed of wind	V	$msec^{-1}$
speed of sound	V_0	$msec^{-1}$
density of air	ρ	kgm^{-3}
viscosity	μ	$kgsec^{-1}m^{-1}$
lift	l	$kgsec^{-2}m$

Here m denotes meters, $\sec$ refers to seconds and kg refers to kilograms. All of these are likely familiar except for μ, which we will discuss in further detail now.

Viscosity is a measure of how much internal friction is experienced when the fluid moves. It is roughly a measure of how “sticky" the fluid is. Consider a piece of area parallel to the direction of motion of the fluid. To say that the viscosity is large is to say that the tangential force applied to this area must be large in order to achieve a given change in speed of the fluid in a direction normal to the tangential force. Thus

$$\mu\,(\text{area})\,(\text{velocity gradient}) = \text{ tangential force}$$

Hence

$$(\text{units on } \mu)\, m^2 \left(\frac{m}{\sec m}\right) = kg\sec^{-2} m$$

Thus the units on μ are

$$kg\sec^{-1} m^{-1}$$

as claimed above.

Returning to our original discussion, you may think that we would want

$$l = f\,(A, B, \theta, V, V_0, \rho, \mu)$$

This is very cumbersome because it depends on seven variables. Also, it is likely that without much care, a change in the units such as going from meters to feet would result in an incorrect value for l. The way to get around this problem is to look for l as a function of dimensionless variables multiplied by something which has units of force. It is helpful because first of all, you will likely have fewer independent variables and secondly, you could expect the formula to hold independent of the way of specifying length, mass and so forth. One looks for

$$l = f\,(g_1, \cdots, g_k)\,\rho V^2 AB$$

where the units on $\rho V^2 AB$ are

$$\frac{kg}{m^3}\left(\frac{m}{\sec}\right)^2 m^2 = \frac{kg \times m}{\sec^2}$$

which are the units of force. Each of these g_i is of the form

$$A^{x_1}B^{x_2}\theta^{x_3}V^{x_4}V_0^{x_5}\rho^{x_6}\mu^{x_7} \tag{1.5.6}$$

and each g_i is independent of the dimensions. That is, this expression must not depend on meters, kilograms, seconds, etc. Thus, placing in the units for each of these quantities, one needs

$$m^{x_1}m^{x_2}\left(m^{x_4}\sec^{-x_4}\right)\left(m^{x_5}\sec^{-x_5}\right)\left(kgm^{-3}\right)^{x_6}\left(kg\sec^{-1}m^{-1}\right)^{x_7} = m^0kg^0\sec^0$$

Notice that there are no units on θ because it is just the radian measure of an angle. Hence its dimensions consist of length divided by length, thus it is dimensionless. Then this leads to the following equations for the x_i.

$$\begin{aligned} m: &\quad x_1 + x_2 + x_4 + x_5 - 3x_6 - x_7 = 0 \\ \sec: &\quad -x_4 - x_5 - x_7 = 0 \\ kg: &\quad x_6 + x_7 = 0 \end{aligned}$$

The augmented matrix for this system is

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

The reduced row-echelon form is given by

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

and so the solutions are of the form

$$\begin{aligned} x_1 &= -x_2 - x_7 \\ x_3 &= x_3 \\ x_4 &= -x_5 - x_7 \\ x_6 &= -x_7 \end{aligned} \tag{1.5.7}$$

Thus, in terms of vectors, the solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} -x_2 - x_7 \\ x_2 \\ x_3 \\ -x_5 - x_7 \\ x_5 \\ -x_7 \\ x_7 \end{bmatrix}$$

Thus the free variables are x_2, x_3, x_5, x_7. By assigning values to these, we can obtain dimensionless variables by placing the values obtained for the x_i in the formula (1.5.6). For example, let $x_2 = 1$ and all the rest of the free variables are 0. This yields

$$x_1 = -1, x_2 = 1, x_3 = 0, x_4 = 0, x_5 = 0, x_6 = 0, x_7 = 0$$

The dimensionless variable is then $A^{-1}B^1$. This is the ratio between the span and the chord. It is called the aspect ratio, denoted as AR. Next let $x_3 = 1$ and all others equal zero. This gives for a dimensionless quantity the angle θ. Next let $x_5 = 1$ and all others equal zero. This gives

$$x_1 = 0, x_2 = 0, x_3 = 0, x_4 = -1, x_5 = 1, x_6 = 0, x_7 = 0$$

Then the dimensionless variable is $V^{-1}V_0^1$. However, it is written as V/V_0. This is called the Mach number $\mathcal{M}$.

Finally, let $x_7 = 1$ and all the other free variables equal 0. Then

$$x_1 = -1, x_2 = 0, x_3 = 0, x_4 = -1, x_5 = 0, x_6 = -1, x_7 = 1$$

then the dimensionless variable which results from this is $A^{-1}V^{-1}\rho^{-1}\mu$. It is customary to write it as $Re = (AV\rho)/\mu$. This one is called the Reynold's number. It is the one which involves viscosity. Thus we would look for

$$l = f(Re, AR, \theta, \mathcal{M})\, kg \times m/\sec^2$$

This is quite interesting because it is easy to vary Re by simply adjusting the velocity or A but it is hard to vary things like μ or ρ. Note that all the quantities are easy to adjust. Now this could be used, along with wind tunnel experiments to get a formula for the lift which would be reasonable. You could also consider more variables and more complicated situations in the same way.

Exercises

Exercise 1

Consider the following diagram of three circuits.

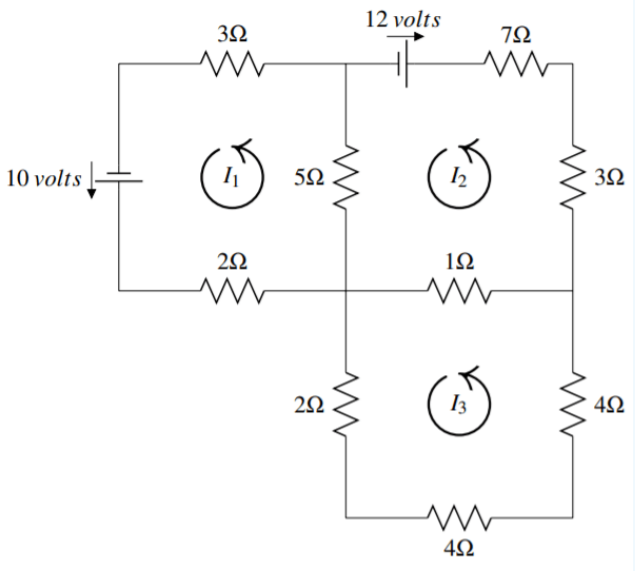


Figure 1.5.2: A circuit with multiple interconnected components, including resistors, voltage sources, and current loops. The circuit is composed of 3 squares: Square 1 on the top left, Square 2 on the top right and Square 3 on the bottom right. Square 1 has counterclockwise current flow I_1, with a 10V voltage source with arrow pointing down, and resistors 2Ω on the bottom, 5Ω on the right, shared with Square 2, and 3Ω above. Square 2 has counterclockwise current flow I_3, with resistors 5Ω on the left shared with Square 1, 1Ω on the bottom shared with Square 3, 3Ω on the right, 7Ω above, and a 12V voltage source with arrow pointing right. Square 3 has counterclockwise current flow I_3, with resistors 2Ω on the left, 4Ω below, 2Ω on the right and 1Ω above shared with Square 2.

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 available at the end of this section.

Exercise 2

Consider a circuit with the following description:

- Left Loop:
  - A 10 V battery
  - A 2 Ω resistor
  - A shared 3 Ω resistor
- Right Loop:
  - A 5 V battery
  - A 4 Ω resistor
  - The same shared 3 Ω resistor (which lies between the two loops)

Suppose that:

- i_1 is the clockwise loop current in the left loop.
- i_2 is the clockwise loop current in the right loop.
- The current through the shared resistor is given by the difference $(i_1 - i_2)$

Write the equations for each loop and then solve for i_1 and i_2.

Exercise 3

Consider a circuit with three loops connected through shared resistors with the following description:

- Loop 1:
  - A 12 V battery
  - A 2 Ω resistor
  - A shared 4 Ω resistor with Loop 2
- Loop 2:
  - A 9 V battery
  - A 3 Ω resistor
  - A shared 4 Ω resistor with Loop 1
  - A shared 5 Ω resistor with Loop 3
- Loop 3:
  - A 6 V battery
  - A 2 Ω resistor
  - A shared 5 Ω resistor with Loop 2

Currents:

- i_1 flows clockwise in Loop 1
- i_2 flows clockwise in Loop 2
- i_3 flows clockwise in Loop 3

The shared resistors carry the current differences between loops:

- 4 Ω resistor (Loops 1 & 2): current is $i_1 - i_2$
- 5 Ω resistor (Loops 2 & 3): current is $i_2 - i_3$

Write the equations for each loop and then solve for i_1, i_2 and i_3.

Answers

Exercise 1

You have

$$\begin{aligned} 2I_1 + 5I_1 + 3I_1 - 5I_2 &= 10 \\ I_2 - I_3 + 3I_2 + 7I_2 + 5I_2 - 5I_1 &= -12 \\ 2I_3 + 4I_3 + 4I_3 + I_3 - I_2 &= 0 \end{aligned} \tag{1.5.8}$$

Simplifying this yields

$$\begin{aligned} 10I_1 - 5I_2 &= 10 \\ -5I_1 + 16I_2 - I_3 &= -12 \\ -I_2 + 11I_3 &= 0 \end{aligned} \tag{1.5.9}$$

The solution is given by

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

This page titled 1.5: Applications to Physics and Resistor Networks is shared under a license and was authored, remixed, and/or curated by LibreTexts.

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

1.6: Chapters Exercises

Exercises

Exercise 1.6.1

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 available at the end of this section.

Exercise 1.6.2

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 available at the end of this section.

Exercise 1.6.3

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 available at the end of this section.

Exercise 1.6.4

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 available at the end of this section.

Exercise 1.6.5

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 available at the end of this section.

Exercise 1.6.6

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

$$\begin{bmatrix} \blacksquare & * & * & * & * & * \\ 0 & \blacksquare & * & * & 0 & * \\ 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * \end{bmatrix}$$

Answer available at the end of this section.

Exercise 1.6.7

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

$$\begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & \blacksquare & * \end{bmatrix}$$

Answer available at the end of this section.

Exercise 1.6.8

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

$$\begin{bmatrix} \blacksquare & * & * & * & * & * \\ 0 & \blacksquare & 0 & * & 0 & * \\ 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * \end{bmatrix}$$

Exercise 1.6.9

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

$$\begin{bmatrix} \blacksquare & * & * & * & * & * \\ 0 & \blacksquare & * & * & 0 & * \\ 0 & 0 & 0 & 0 & \blacksquare & 0 \\ 0 & 0 & 0 & 0 & * & \blacksquare \end{bmatrix}$$

Answer available at the end of this section.

Exercise 1.6.10

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 available at the end of this section.

Exercise 1.6.11

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 available at the end of this section.

Exercise 1.6.12

Find h such that

$$\begin{bmatrix} 2 & h & 4 \\ 3 & 6 & 7 \end{bmatrix}$$

is the augmented matrix of an *inconsistent* system.

Answer available at the end of this section.

Exercise 1.6.13

Find h such that

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

is the augmented matrix of a *consistent* system.

Answer available at the end of this section.

Exercise 1.6.14

Find h such that

$$\begin{bmatrix} 1 & 1 & 4 \\ 3 & h & 12 \end{bmatrix}$$

is the augmented matrix of a *consistent* system.

Answer available at the end of this section.

Exercise 1.6.15

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

$$\begin{bmatrix} 1 & h & 2 \\ 2 & 4 & k \end{bmatrix}$$

Answer available at the end of this section.

Exercise 1.6.16

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

$$\begin{bmatrix} 1 & 2 & 2 \\ 2 & h & k \end{bmatrix}$$

Answer available at the end of this section.

Exercise 1.6.17

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 available at the end of this section.

Exercise 1.6.18

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 available at the end of this section.

Exercise 1.6.19

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 available at the end of this section.

Exercise 1.6.20

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 1.6.21

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 1.6.22

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 1.6.23

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 1.6.24

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 1.6.25

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 1.6.26

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 1.6.27

Find the solution of the system whose augmented matrix is

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

Exercise 1.6.28

Find the solution of the system whose augmented matrix is

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

Answer available at the end of this section.

Exercise 1.6.29

Find the solution of the system whose augmented matrix is

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

Answer available at the end of this section.

Exercise 1.6.30

Find the solution of the system whose augmented matrix is

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

Answer available at the end of this section.

Exercise 1.6.31

Find the solution of the system whose augmented matrix is

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

Answer available at the end of this section.

Exercise 1.6.32

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

Answer available at the end of this section.

Exercise 1.6.33

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

Answer available at the end of this section.

Exercise 1.6.34

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

Answer available at the end of this section.

Exercise 1.6.35

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

Answer available at the end of this section.

Exercise 1.6.36

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

Answer available at the end of this section.

Exercise 1.6.37

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

Answer available at the end of this section.

Exercise 1.6.38

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

Answer available at the end of this section.

Exercise 1.6.39

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

Answer available at the end of this section.

Exercise 1.6.40

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

Answer available at the end of this section.

Exercise 1.6.41

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 available at the end of this section.

Exercise 1.6.42

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 available at the end of this section.

Exercise 1.6.43

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 available at the end of this section.

Exercise 1.6.44

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 available at the end of this section.

Exercise 1.6.45

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

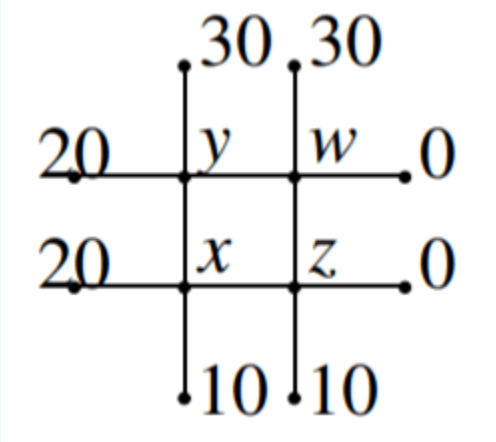


Figure 1.6.1

Answer available at the end of this section.

Exercise 1.6.46

Find the rank of the following matrix.

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

Exercise 1.6.47

Find the rank of the following matrix.

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

Exercise 1.6.48

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 1.6.49

Find the rank of the following matrix.

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

Exercise 1.6.50

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 1.6.51

Find the rank of the following matrix.

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

Exercise 1.6.52

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 1.6.53

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 1.6.54

Find the rank of the following matrix.

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

Exercise 1.6.55

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 1.6.56

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 1.6.57

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

Answer available at the end of this section.

Exercise 1.6.58

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 available at the end of this section.

Exercise 1.6.59

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 available at the end of this section.

Exercise 1.6.60

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. $CaCl_2 + Na_3PO_4 \rightarrow Ca_3(PO_4)_2 + NaCl$

Answer available at the end of this section.

Exercise 1.6.61

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 1.6.62

Consider the following diagram of four circuits.

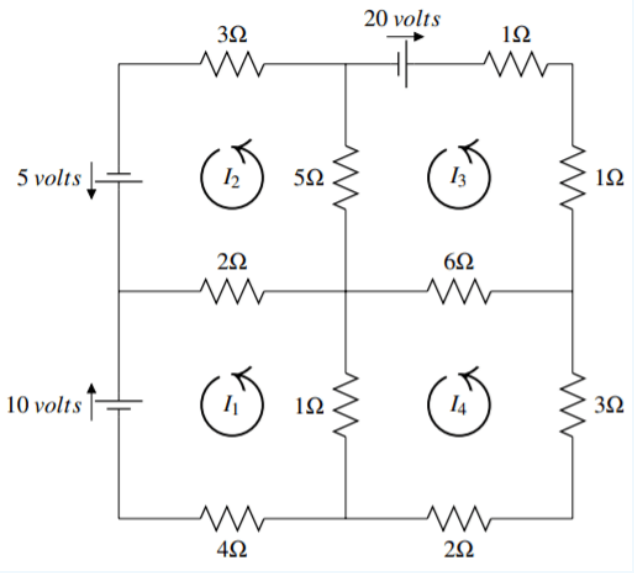


Figure 1.6.2

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 available at the end of this section.

Exercise 1.6.63

Consider the following diagram of three circuits.

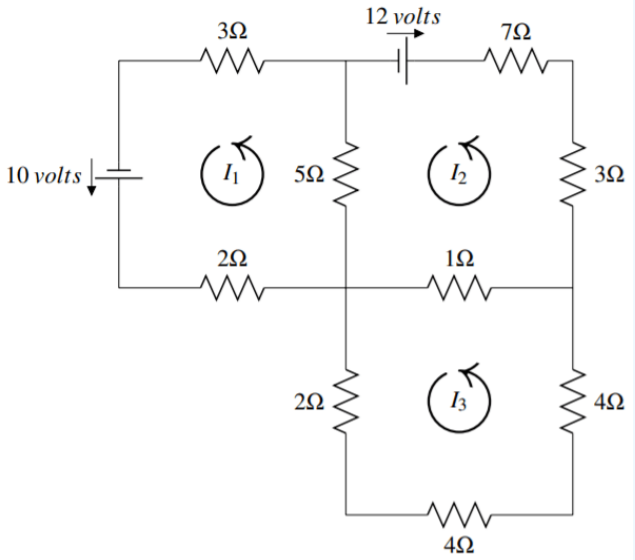


Figure 1.6.3

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 available at the end of this section.

Answers

Exercise 1.6.1

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

Exercise 1.6.2

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

Exercise 1.6.3

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

Exercise 1.6.4

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 1.6.5

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

Exercise 1.6.6

The solution exists but is not unique.

Exercise 1.6.7

A solution exists and is unique.

Exercise 1.6.9

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

Exercise 1.6.10

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

Exercise 1.6.11

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

Exercise 1.6.12

$h=4$

Exercise 1.6.13

Any h will work.

Exercise 1.6.14

Any h will work.

Exercise 1.6.15

If $h \neq 2$ there will be a unique solution for any k. If $h=2$ and $k \neq 4$, there are no solutions. If $h=2$ and $k=4$, then there are infinitely many solutions.

Exercise 1.6.16

If $h \neq 4$, then there is exactly one solution. If $h=4$ and $k \neq 4$, then there are no solutions. If $h=4$ and $k=4$, then there are infinitely many solutions.

Exercise 1.6.17

There is no solution. The system is inconsistent. You can see this from the augmented matrix. $\left[\begin{array}{rrrrr} 1 & 2 & 1 & -1 & 2 \\ 1 & -1 & 1 & 1 & 1 \\ 2 & 1 & -1 & 0 & 1 \\ 4 & 2 & 1 & 0 & 5 \end{array}\right]$, reduced row-echelon form: $\left[\begin{array}{rrrrr} 1 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 1 & 0 & -\frac{2}{3} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]$.

Exercise 1.6.18

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

Exercise 1.6.19

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

Exercise 1.6.28

The reduced row-echelon form is $\left[\begin{array}{rrrr} 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & -\frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 0 \end{array}\right]$. 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 1.6.29

The reduced row-echelon form is $\left[\begin{array}{rrrr} 1 & 0 & 4 & 2 \\ 0 & 1 & -4 & -1 \end{array}\right]$ and so the solution is $z=t, y=4t, x=2-4t$.

Exercise 1.6.30

The reduced row-echelon form is $\left[\begin{array}{rrrrrr} 1 & 0 & 0 & 0 & 9 & 3 \\ 0 & 1 & 0 & 0 & -4 & 0 \\ 0 & 0 & 1 & 0 & -7 & -1 \\ 0 & 0 & 0 & 1 & 6 & 1 \end{array}\right]$ and so $x_5=t, x_4=1-6t, x_3=-1+7t, x_2=4t, x_1=3-9t$.

Exercise 1.6.31

The reduced row-echelon form is $\left[\begin{array}{rrrrrr} 1 & 0 & 2 & 0 & -\frac{1}{2} & \frac{5}{2} \\ 0 & 1 & 0 & 0 & \frac{1}{2} & \frac{3}{2} \\ 0 & 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$. 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 1.6.32

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

Exercise 1.6.33

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

Exercise 1.6.34

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

Exercise 1.6.35

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

Exercise 1.6.36

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

Exercise 1.6.37

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

Exercise 1.6.38

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

Exercise 1.6.39

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

Exercise 1.6.40

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

Exercise 1.6.41

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

Exercise 1.6.42

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

Exercise 1.6.43

Yes. It has a unique solution.

Exercise 1.6.44

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

Exercise 1.6.45

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

Exercise 1.6.57

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 1.6.58

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 1.6.59

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

Exercise 1.6.60

Exercise 1.6.62

The other two equations are

$$\begin{array}{c} 6I_3-6I_4+I_3+I_3+5I_3-5I_2=-20 \\ 2I_4+3I_4+6I_4-6I_3+I_4-I_1=0 \end{array} \tag{1.6.1}$$

Then the system is

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

The solution is:

$$\begin{array}{l} I_1=-\frac{750}{373} \\ I_2=-\frac{1421}{1119} \\ I_3=-\frac{3061}{1119} \\ I_4=-\frac{1718}{1119} \end{array} \tag{1.6.2}$$

Exercise 1.6.63

You have

$$\begin{array}{c} 2I_1+5I_1+3I_1-5I_2=10 \\ I_2-I_3+3I_2+7I_2+5I_2-5I_1=-12 \\ 2I_3+4I_3+4I_3+I_3-I_2=0 \end{array} \tag{1.6.3}$$

Simplifying this yields

$$\begin{array}{c} 10I_1-5I_2=10 \\ -5I_1+16I_2-I_3=-12 \\ -I_2+11I_3=0 \end{array} \tag{1.6.4}$$

The solution is given by

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

This page titled 1.6: Chapters Exercises is shared under a license and was authored, remixed, and/or curated by LibreTexts.

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

2: MATRICES

2.1: Matrix Addition, Scalar Multiplication, and Transposition
2.2: Matrix Multiplication
2.3: The Identity and Inverses
2.4: Finding the Inverse of a Matrix
2.5: Elementary Matrices
2.6: LU Factorization

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

2.1: Matrix Addition, Scalar Multiplication, and Transposition

Learning Objectives

- Add two matrices of the same size entry by entry.
- Multiply a matrix by a scalar and describe the effect on its entries.
- Recognize when matrix addition and scalar multiplication are not defined.
- Apply the properties of matrix addition, such as commutativity and associativity.
- Use the distributive properties of scalar multiplication with respect to matrix addition.
- Represent systems of equations and other data using matrix addition and scalar multiplication

Definition: A Matrix

A **matrix** is a rectangular array of numbers arranged in rows and columns. Matrices are usually denoted by uppercase letters: $A, B, C, \ldots$. If a matrix has m rows and n columns, its **size** is said to be $m \times n$. The individual numbers inside a matrix are called **entries**. Each entry is written using a lowercase letter with subscripts that show its position in the matrix. The entry in row i and column j is called the (i, j)-entry and is denoted by a_{ij}.

$$A = [a_{ij}]_{m \times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{2.1.1}$$

Example

Given the following matrixes:

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

1- Find the sizes of the following matrices.

2- Find values for the following entries in the above matrices if exist:

- a_{12}
- a_{21}
- b_{22}
- c_{31}.

Solution

1- Since the matrix A shown has 2 rows and 3 columns. In general, a matrix with m rows and n columns is referred to as an $m \times n$ ***matrix*** or as having **size** $m \times n$. Thus matrices A, B, and C above have sizes 2×3, 2×2, and 3×1, respectively.

2- Here are the values of the given entries:

- a_{12}: The first index 1 means row 1, and the second index 2 means column 2. In matrix A, this is 2. So $a_{12} = 2$.

- a_{21}: The first index 2 means row 2, and the second index 1 means column 1. In matrix A, this is 0. So $a_{21} = 0$.
- b_{22}: The first index 2 means row 2, and the second index 2 means column 2. In matrix B, this is 2. So $b_{22} = 2$.
- c_{31}: The first index 3 refers to row 3 and the second index 1 to column 1. In matrix C, this is 2. So $c_{31} = 2$.
- The entry c_{13} does not exist. Matrix C has only one column of size 3×1, which means it has 3 rows and only 1 column. The index c_{13} refers to the entry in row 1, column 3. Since there is no third column in C, this entry cannot be found.

Definition: Equal Matrices

Two matrices A and B are called **equal** (written $A = B$) if and only if:

1. *They have the same size.*
2. *Corresponding entries are equal.*

Two points (x_1, y_1) and (x_2, y_2) in the plane are equal if and only if they have the same coordinates; that is, $(x_1 = x_2)$ and $(y_1 = y_2)$.

Similarly, two matrices are equal if and only if they have the same size and all their corresponding entries are equal.

Example 2.1.1

Given $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 & -1 \\ 3 & 0 & 1 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$ discuss the possibility that $A = B$, $B = C$, $A = C$.

Solution

$A = B$ is impossible because A and B are of different sizes: A is 2×2 whereas B is 2×3. Similarly, $B = C$ is impossible. But $A = C$ is possible provided that corresponding entries are equal: $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$ means $a = 1$, $b = 0$, $c = -1$, and $d = 2$.

After understanding the general definition of a matrix, it is useful to classify matrices into special categories based on their shape or the arrangement of their entries. These classifications help us recognize important structures that appear frequently in applications and make certain computations easier. Let us look at some common \textbf{types of matrices}.

Type	Definition	Example
Row Matrix	A matrix with only one row (*1 × n*).	$\begin{bmatrix} 2 & -1 & 5 & 0 \end{bmatrix}$
Column Matrix	A matrix with only one column (*m × 1*).	$\begin{bmatrix} 3 \\ -2 \\ 4 \end{bmatrix}$
Square Matrix	A matrix with the same number of rows and columns (*n × n*).	$\begin{bmatrix} 1 & 2 & -11 \\ 3 & 4 & 0 \\ -4 & 0 & 0.05 \end{bmatrix}$

Zero Matrix	A matrix in which all entries are zero.	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
Diagonal Matrix	A square matrix with all non-diagonal entries equal to zero.	$\begin{bmatrix} 5 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$
Identity Matrix	A square matrix with 1's on the main diagonal and 0's elsewhere.	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Symmetric Matrix	A square matrix that is equal to its transpose, i.e. entries are mirrored across the diagonal.	$\begin{bmatrix} 2 & 3 & 1 \\ 3 & 4 & -2 \\ 1 & -2 & 5 \end{bmatrix}$
Triangular Matrix	A square matrix is called upper triangular if all the entries below the main diagonal are zero. It is called lower triangular if all the entries above the main diagonal are zero.	Upper Triangle: $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$ Lower Triangle: $\begin{bmatrix} 7 & 0 & 0 \\ 8 & 9 & 0 \\ 1 & 4 & 2 \end{bmatrix}$

Example

Identify the type of each of the following matrices

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{2.1.11}$$

Solution

- Matrix A is a square matrix of size 2×2. It is also a zero matrix, since all entries are zero, and it can be considered a diagonal matrix because all off-diagonal entries are zero.
- Matrix B is a square matrix of size 3×3. It is diagonal, since only the main diagonal entries are nonzero, and it is also symmetric.
- Matrix C is a square matrix of size 3×3. It is diagonal, and at the same time both upper triangular and lower triangular, since all non-diagonal entries are zero.

Now that we know when two matrices are considered equal, we can move on to operations that combine or modify them. Just as we add or scale numbers, we can also add matrices and multiply a matrix by a scalar (a single number). However, with matrices there are specific constraints that we need to investigate. These operations will allow us to manipulate matrices in useful ways and prepare us for more advanced topics such as matrix multiplication and linear transformations.

Matrix Addition

Definition: Matrix Addition

*If A and B are matrices of the same size, their **sum** $A + B$ is the matrix formed by adding corresponding entries.*

If $A = [a_{ij}]$ and $B = [b_{ij}]$, this takes the form

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

Note that addition is *not* defined for matrices of different sizes.

Example 2.1.2

If $A = \begin{bmatrix} 2 & 1 & 3 \\ -1 & 2 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & 6 \end{bmatrix}$, compute $A + B$.

Solution

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

Example 2.1.3

Find a, b, and c if $\begin{bmatrix} a & b & c \end{bmatrix} + \begin{bmatrix} c & a & b \end{bmatrix} = \begin{bmatrix} 3 & 2 & -1 \end{bmatrix}$.

Solution

Add the matrices on the left side to obtain

$$\begin{bmatrix} a+c & b+a & c+b \end{bmatrix} = \begin{bmatrix} 3 & 2 & -1 \end{bmatrix}$$

Because corresponding entries must be equal, this gives three equations: $a + c = 3$, $b + a = 2$, and $c + b = -1$. Solving these yields $a = 3$, $b = -1$, $c = 0$.

Proposition: Properties of Matrix Addition

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

- Commutative Law of Addition

$$A + B = B + A \tag{2.1.12}$$

- Associative Law of Addition

$$(A + B) + C = A + (B + C) \tag{2.1.13}$$

- Existence of an Additive Identity

$$\text{There exists a zero matrix } 0 \text{ such that } A + 0 = A \tag{2.1.14}$$

- Existence of an Additive Inverse

$$\text{There exists a matrix } -A \text{ such that } A + (-A) = 0 \tag{2.1.15}$$

Answer available at the end of this section.

We call the zero matrix in (2.1.14) the **additive identity**. Similarly, we call the matrix $-A$ in (2.1.15) the **additive inverse**. $-A$ is defined to equal $(-1)\, A = [-a_{ij}]$. In other words, every entry of A is multiplied by -1, we will discuss in scalar multiplication. It si not hard to prove the above proberties as they are derived from the addition Properties of real numbers.

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: Scalar Multiplication of Matrices

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

Consider the following example.

Example: 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 the scalar multiplication definition, 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
$$1A = A$$

Example 2.1.4

Let $A = \begin{bmatrix} 3 & -1 & 0 \\ 1 & 2 & -4 \end{bmatrix}$, $B = \begin{bmatrix} 1 & -1 & 1 \\ -2 & 0 & 6 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 & -2 \\ 3 & 1 & 1 \end{bmatrix}$. Compute $-A$, $A-B$, and $A+B-C$.

Solution

$$-A = \begin{bmatrix} -3 & 1 & 0 \\ -1 & -2 & 4 \end{bmatrix}$$
$$A-B = \begin{bmatrix} 3-1 & -1-(-1) & 0-1 \\ 1-(-2) & 2-0 & -4-6 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -1 \\ 3 & 2 & -10 \end{bmatrix}$$
$$A+B-C = \begin{bmatrix} 3+1-1 & -1-1-0 & 0+1-(-2) \\ 1-2-3 & 2+0-1 & -4+6-1 \end{bmatrix} = \begin{bmatrix} 3 & -2 & 3 \\ -4 & 1 & 1 \end{bmatrix}$$

Example 2.1.5

Solve $\begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix} + X = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$ where X is a matrix.

Solution

We solve a numerical equation $a + x = b$ by subtracting the number a from both sides to obtain $x = b - a$. This also works for matrices. To solve $\begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix} + X = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$ simply subtract the matrix $\begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix}$ from both sides to get

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

The reader should verify that this matrix X does indeed satisfy the original equation.

Example 2.1.6

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

Solution

$$5A = \begin{bmatrix} 15 & -5 & 20 \\ 10 & 0 & 30 \end{bmatrix}, \quad \frac{1}{2}B = \begin{bmatrix} \frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & \frac{3}{2} & 1 \end{bmatrix}$$
$$3A - 2B = \begin{bmatrix} 9 & -3 & 12 \\ 6 & 0 & 18 \end{bmatrix} - \begin{bmatrix} 2 & 4 & -2 \\ 0 & 6 & 4 \end{bmatrix} = \begin{bmatrix} 7 & -7 & 14 \\ 6 & -6 & 14 \end{bmatrix}$$

If A is any matrix, note that kA is the same size as A for all scalars k. We also have

$$0A = 0 \quad \text{and} \quad k0 = 0$$

because the zero matrix has every entry zero. In other words, $kA = 0$ if either $k = 0$ or $A = 0$. The converse of this statement is also true, as Example [exa:002159] shows.

Example 2.1.7

If $kA = 0$, show that either $k = 0$ or $A = 0$.

Solution

Write $A = [a_{ij}]$ so that $kA = 0$ means $ka_{ij} = 0$ for all i and j. If $k = 0$, there is nothing to do. If $k \neq 0$, then $ka_{ij} = 0$ implies that $a_{ij} = 0$ for all i and j; that is, $A = 0$.

Example 2.1.8

Simplify $2(A + 3C) - 3(2C - B) - 3[2(2A + B - 4C) - 4(A - 2C)]$ where A, B, and C are all matrices of the same size.

Solution

The reduction proceeds as though A, B, and C were variables.

$$\begin{aligned} &2(A + 3C) - 3(2C - B) - 3[2(2A + B - 4C) - 4(A - 2C)] \\ &\quad = 2A + 6C - 6C + 3B - 3[4A + 2B - 8C - 4A + 8C] \\ &\quad = 2A + 3B - 3[2B] \\ &\quad = 2A - 3B \end{aligned}$$

So far, we have learned how to add matrices and multiply them by scalars, which allow us to combine and scale information within a matrix. Another important operation is the \textbf{transpose} of a matrix. This operation does not change the numerical values of the entries but instead rearranges their positions, giving us a new perspective on the same data. Understanding the transpose will be essential when we discuss symmetric matrices, inner products, and many applications in linear algebra.

Transpose of a Matrix

Many results about a matrix A involve the *rows* of A, and the corresponding result for columns is derived in an analogous way, essentially by replacing the word *row* by the word *column* throughout. The following definition is made with such applications in

mind.

Definition: Transpose of a Matrix

*If A is an $m \times n$ matrix, the **transpose** of A, written A^T, is the $n \times m$ matrix whose rows are just the columns of A in the same order.*

In other words, the first row of A^T is formed from the first column of A (with the entries listed in order). Similarly, the second row of A^T comes from the second column of A, and so on for the remaining rows and columns.

Formally, if $A = [a_{ij}]$ is a matrix and we write $A^T = [b_{ij}]$, then the entry b_{ij} in A^T is equal to the entry a_{ji} in A. In other words,

$$b_{ij} = a_{ji}. \tag{2.1.16}$$

Example 2.1.9

Write down the transpose of each of the following matrices.

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

Solution

$$A^T = \begin{bmatrix} 1 & 3 & 2 \end{bmatrix}, \; B^T = \begin{bmatrix} 5 \\ 2 \\ 6 \end{bmatrix}, \; C^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}, \text{ and } D^T = D.$$

$$\text{If } A = [a_{ij}], \text{ then } A^T = [a_{ji}]. \tag{2.1.17}$$

This is useful in verifying the following properties of transposition.

Theorem 2.1.1

Let A and B denote matrices of the same size, and let k denote a scalar.

1. *If A is an $m \times n$ matrix, then A^T is an $n \times m$ matrix.*
2. $(A^T)^T = A$.
3. $(kA)^T = kA^T$.
4. $(A + B)^T = A^T + B^T$.

We can use the definition of transpose and the properties of matrix addition and scalar multiplication to verify the above theorem. Here is a example.

Example 2.1.1

Consider the matrix A, and let k denote a scalar. Verify the following equation.

$(kA)^T = kA^T$.

Solution

Proof. Property 1 is part of the definition of A^T, and Property 2 follows from ([eq:transpose]). As to Property 3: If $A = [a_{ij}]$, then $kA = [ka_{ij}]$, so ([eq:transpose]) gives

$$(kA)^T = [ka_{ji}] = k[a_{ji}] = kA^T$$

Finally, if $B = [b_{ij}]$, then $A + B = [c_{ij}]$ where $c_{ij} = a_{ij} + b_{ij}$ Then ([eq:transpose]) gives Property 4:

$$(A + B)^T = [c_{ij}]^T = [c_{ji}] = [a_{ji} + b_{ji}] = [a_{ji}] + [b_{ji}] = A^T + B^T$$

There is another useful way to think of transposition. If $A = [a_{ij}]$ is an $m \times n$ matrix, the elements $a_{11}, a_{22}, a_{33}, \ldots$ are called the **main diagonal** of A. Hence the main diagonal extends down and to the right from the upper left corner of the matrix A; shown in bold in the following examples:

$$\begin{pmatrix} \mathbf{a_{11}} & a_{12} \\ a_{21} & \mathbf{a_{22}} \\ a_{31} & a_{32} \end{pmatrix} \begin{pmatrix} \mathbf{a_{11}} & a_{12} & a_{13} \\ a_{21} & \mathbf{a_{22}} & a_{23} \end{pmatrix} \begin{pmatrix} \mathbf{a_{11}} & a_{12} & a_{13} \\ a_{21} & \mathbf{a_{22}} & a_{23} \\ a_{31} & a_{32} & \mathbf{a_{33}} \end{pmatrix} \begin{pmatrix} \mathbf{a_{11}} \\ a_{21} \end{pmatrix} \tag{2.1.18}$$

Thus forming the transpose of a matrix A can be viewed as "flipping" A about its main diagonal, or as "rotating" A through 180° about the line containing the main diagonal. This makes Property 2 in Theorem 2.1.2 transparent.

Example 2.1.10

Solve for A if $\left(2A^T - 3\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}\right)^T = \begin{bmatrix} 2 & 3 \\ -1 & 2 \end{bmatrix}$.

Solution

Using Theorem 2.1.2, the left side of the equation is

$$\left(2A^T - 3\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}\right)^T = 2\left(A^T\right)^T - 3\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}^T = 2A - 3\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}$$

Hence the equation becomes

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

Thus $2A = \begin{bmatrix} 2 & 3 \\ -1 & 2 \end{bmatrix} + 3\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 5 & 5 \end{bmatrix}$, so finally $A = \frac{1}{2}\begin{bmatrix} 5 & 0 \\ 5 & 5 \end{bmatrix} = \frac{5}{2}\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

In solving problems, the transpose often appears as a useful tool. For example, it plays an important role when checking whether a matrix is symmetric. The transpose is an operation, and we describe it by saying we "take the transpose of A."

Reacll that a matrix A is called **symmetric** if $A = A^T$. A symmetric matrix A is necessarily square (if A is $m \times n$, then A^T is $n \times m$, so $A = A^T$ forces $n = m$). The name comes from the fact that these matrices exhibit a symmetry about the main diagonal. That is, entries that are directly across the main diagonal from each other are equal.

For example, $\begin{bmatrix} a & b & c \\ b' & d & e \\ c' & e' & f \end{bmatrix}$ is symmetric when $b = b'$, $c = c'$, and $e = e'$.

Example 2.1.11

If A and B are symmetric $n \times n$ matrices, show that $A + B$ is symmetric.

Solution

We have $A^T = A$ and $B^T = B$, so, by Theorem [thm:002240], we have $(A+B)^T = A^T + B^T = A + B$. Hence $A + B$ is symmetric.

Example 2.1.12

Suppose a square matrix A satisfies $A = 2A^T$. Show that necessarily $A = 0$.

Solution

If we iterate the given equation, Theorem [thm:002240] gives

$$A = 2A^T = 2\left[2A^T\right]^T = 2\left[2(A^T)^T\right] = 4A$$

Subtracting A from both sides gives $3A = 0$, so $A = \frac{1}{3}(0) = 0$.

---

1. If p and q are statements, we say that p implies q if q is true whenever p is true. Then "p if and only if q" means that both p implies q and q implies p. See Appendix [chap:appbproofs] for more on this.

Exercises

Exercise

Find a, b, c, and d if

a. $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} c-3d & -d \\ 2a+d & a+b \end{bmatrix}$

b. $\begin{bmatrix} a-b & b-c \\ c-d & d-a \end{bmatrix} = 2\begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix}$

c. $3\begin{bmatrix} a \\ b \end{bmatrix} + 2\begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

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

Answer available at the end of this section.

Exercise

Compute the following:

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

b. $3\begin{bmatrix} 3 \\ -1 \end{bmatrix} - 5\begin{bmatrix} 6 \\ 2 \end{bmatrix} + 7\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

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

d. $\begin{bmatrix} 3 & -1 & 2 \end{bmatrix} - 2\begin{bmatrix} 9 & 3 & 4 \end{bmatrix} + \begin{bmatrix} 3 & 11 & -6 \end{bmatrix}$

e. $\begin{bmatrix} 1 & -5 & 4 & 0 \\ 2 & 1 & 0 & 6 \end{bmatrix}^T$

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

g. $\begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix} - 2\begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix}^T$

h. $3\begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}^T - 2\begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}$

Answer available at the end of this section.

Exercise

Let $A = \begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix}$,
$B = \begin{bmatrix} 3 & -1 & 2 \\ 0 & 1 & 4 \end{bmatrix}$, $C = \begin{bmatrix} 3 & -1 \\ 2 & 0 \end{bmatrix}$,
$D = \begin{bmatrix} 1 & 3 \\ -1 & 0 \\ 1 & 4 \end{bmatrix}$, and $E = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$.

Compute the following (where possible).

a. $3A - 2B$
b. $5C$
c. $3E^T$
d. $B + D$
e. $4A^T - 3C$
f. $(A + C)^T$
g. $2B - 3E$
h. $A - D$
i. ((B - 2E)^{T}\)

Answer available at the end of this section.

Exercise

Find A if:

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

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

Answer available at the end of this section.

Exercise

Find A in terms of B if:

a. $A + B = 3A + 2B$
b. $2A - B = 5(A + 2B)$

Answer available at the end of this section.

Exercise

If X, Y, A, and B are matrices of the same size, solve the following systems of equations to obtain X and Y in terms of A and B.

a. $5X + 3Y = A \; 2X + Y = B$
b. $4X + 3Y = A \; 5X + 4Y = B$

Answer available at the end of this section.

Example

Find all matrices X and Y such that:

a. $3X - 2Y = \begin{bmatrix} 3 & -1 \end{bmatrix}$
b. $2X - 5Y = \begin{bmatrix} 1 & 2 \end{bmatrix}$

Solution

$Y = (s, t)$, $X = \frac{1}{2}(1 + 5s, 2 + 5t)$; s and t arbitrary

Exercise

Simplify the following expressions where A, B, and C are matrices.

a. $2\left[9(A - B) + 7(2B - A)\right]$
$-2\left[3(2B + A) - 2(A + 3B) - 5(A + B)\right]$
b. $5\left[3(A - B + 2C) - 2(3C - B) - A\right]$
$+2\left[3(3A - B + C) + 2(B - 2A) - 2C\right]$

Answer available at the end of this section.

Exercise

If A is any 2×2 matrix, show that:

a. $A = a\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ for some numbers a, b, c, and d.
b. $A = p\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + q\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + r\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} + s\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ for some numbers p, q, r, and s.

Answer available at the end of this section.

Exercise

Let $A = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}$,
$B = \begin{bmatrix} 0 & 1 & 2 \end{bmatrix}$, and $C = \begin{bmatrix} 3 & 0 & 1 \end{bmatrix}$. If
$rA + sB + tC = 0$ for some scalars r, s, and t, show that necessarily $r = s = t = 0$.

Answer available at the end of this section.

Exercise

a. If $Q + A = A$ holds for every $m \times n$ matrix A, show that $Q = 0_{mn}$.
b. If A is an $m \times n$ matrix and $A + A' = 0_{mn}$, show that \(A^\prime = -A\

Answer available at the end of this section.

Exercise

If A denotes an $m \times n$ matrix, show that $A = -A$ if and only if $A = 0$.

Exercise \

In each case determine all s and t such that the given matrix is symmetric:

a. $\begin{bmatrix} 1 & s \\ -2 & t \end{bmatrix}$
b. $\begin{bmatrix} s & t \\ st & 1 \end{bmatrix}$
c. $\begin{bmatrix} s & 2s & st \\ t & -1 & s \\ t & s^2 & s \end{bmatrix}$
d. $\begin{bmatrix} 2 & s & t \\ 2s & 0 & s+t \\ 3 & 3 & t \end{bmatrix}$

Answer available at the end of this section.

Exercise

In each case find the matrix A.

a. $\left(A + 3\begin{bmatrix} 1 & -1 & 0 \\ 1 & 2 & 4 \end{bmatrix}\right)^T = \begin{bmatrix} 2 & 1 \\ 0 & 5 \\ 3 & 8 \end{bmatrix}$
b. $\left(3A^T + 2\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right)^T = \begin{bmatrix} 8 & 0 \\ 3 & 1 \end{bmatrix}$
c. $\left(2A - 3\begin{bmatrix} 1 & 2 & 0 \end{bmatrix}\right)^T = 3A^T + \begin{bmatrix} 2 & 1 & -1 \end{bmatrix}^T$
d. $\left(2A^T - 5\begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}\right)^T = 4A - 9\begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}$

Answer available at the end of this section.

Exercise

Let A and B be symmetric (of the same size). Show that each of the following is symmetric.

a. $(A - B)$

b. kA for any scalar k

Answer available at the end of this section.

Exercise

Show that $A + A^T$ is symmetric for *any* square matrix A.

Exercise

If A is a square matrix and $A = kA^T$ where $k \neq \pm 1$, show that $A = 0$.

Exercise

In each case either show that the statement is true or give an example showing it is false.

a. If $A + B = A + C$, then B and C have the same size.
b. If $A + B = 0$, then $B = 0$.
c. If the $(3, 1)$-entry of A is 5, then the $(1, 3)$-entry of A^T is -5.
d. A and A^T have the same main diagonal for every matrix A.
e. If B is symmetric and $A^T = 3B$, then $A = 3B$.
f. If A and B are symmetric, then $kA + mB$ is symmetric for any scalars k and m.

Answer available at the end of this section.

Exercise

A square matrix W is called **skew-symmetric** if $W^T = -W$. Let A be any square matrix.

a. Show that $A - A^T$ is skew-symmetric.
b. Find a symmetric matrix S and a skew-symmetric matrix W such that $A = S + W$.
c. Show that S and W in part (b) are uniquely determined by A.

Answer available at the end of this section.

Exercise

If W is skew-symmetric (Exercise $2.1.20$), show that the entries on the main diagonal are zero.

2.

Exercise

Prove the following parts of Theorem [thm:002170].

a. $(k + p)A = kA + pA$
b. $(kp)A = k(pA)$

Answer available at the end of this section.

1.

Exercise

Let $A, A_1, A_2, \ldots, A_n$ denote matrices of the same size. Use induction on n to verify the following extensions of properties 5 and 6 of Theorem 2.1.1.

a. (k(A_{1} + A_{2} + \dots + A_{n}) = kA_{1} + kA_{2} + \dots + kA_{n}\) for any number k
b. $(k_1 + k_2 + \cdots + k_n)A = k_1A + k_2A + \cdots + k_nA$ for any numbers $k_1, k_2, \ldots, k_n$

Example

Let A be a square matrix. If $A = pB^T$ and $B = qA^T$ for some matrix B and numbers p and q, show that either $A = 0 = B$ or $pq = 1$.

Answers

Proposition: Properties of Matrix Addition

Exercise

b. $(a\ b\ c\ d) = (-2, -4, -6, 0) + t(1, 1, 1, 1)$,
c. t arbitrary $a = b = c = d = t$, t arbitrary

Exercise

b. $\begin{bmatrix} -14 \\ -20 \end{bmatrix} (-12, 4, -12)$
c. $\begin{bmatrix} 0 & 1 & -2 \\ -1 & 0 & 4 \\ 2 & -4 & 0 \end{bmatrix}$
d. $\begin{bmatrix} 4 & -1 \\ -1 & -6 \end{bmatrix}$

Exercise

b. $\begin{bmatrix} 15 & -5 \\ 10 & 0 \end{bmatrix}$
c. Impossible
d. $\begin{bmatrix} 5 & 2 \\ 0 & -1 \end{bmatrix}$
e. Impossible

Exercise

b. $\begin{bmatrix} 2 \\ -\frac{1}{2} \end{bmatrix}$

Exercise

b. $A = -\frac{11}{3}B$

Exercise

b. $X = 4A - 3B$, $Y = 4B - 5A$

Exercise

b. $20A - 7B + 2C$

Exercise

b. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $(p, q, r, s) = \frac{1}{2}(2d, a + b - c - d, a - b + c - d, -a + b + c + d)$.

Exercise

Exercise

b. If $A + A' = 0$ then $-A = -A + 0 = -A + (A + A') = (-A + A) + A' = 0 + A' = A'$
b. $s = 1$ or $t = 0$
c. $s = 0$, and $t = 3$

Exercise

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

c. $\begin{bmatrix} 2 & 7 \\ -\frac{9}{2} & -5 \end{bmatrix}$

Exercise

a. $A = A^T$,
b. $(kA)^T = kA^T = kA$.

Exercise

a. False. Take $B = -A$ for any $A \neq 0$.
b. True. Transposing fixes the main diagonal.
c. True. $(kA + mB)^T = (kA)^T + (mB)^T = kA^T + mB^T = kA + mB$

Exercise

c. Suppose $A = S + W$, where $S = S^T$ and $W = -W^T$. Then $A^T = S^T + W^T = S - W$, so $A + A^T = 2S$ and $A - A^T = 2W$. Hence $S = \frac{1}{2}(A + A^T)$ and $W = \frac{1}{2}(A - A^T)$ are uniquely determined by A.

Exercise

b. If $A = [a_{ij}]$ then $(kp)A = [(kp)a_{ij}] = [k(pa_{ij})] = k[pa_{ij}] = k(pA)$.

This page titled 2.1: Matrix Addition, Scalar Multiplication, and Transposition is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by W. Keith Nicholson via source content that was edited to the style and standards of the LibreTexts platform.

- 2.1: Matrix Addition, Scalar Multiplication, and Transposition by W. Keith Nicholson is licensed CC BY-NC-SA 4.0. Original source: https://lyryx.com/linear-algebra-applications.
- 2.1: Matrix Arithmetic by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

2.2: Matrix Multiplication

Learning Objectives

After studying this section, I can:

- Determine when two matrices are conformable for multiplication.
- Multiply a row vector by a column vector and identify the dot product.
- Multiply a column vector by a row vector and identify the outer product.
- Multiply an $m \times n$ matrix by an $n \times 1$ column vector and interpret the result as a linear combination of the columns of the matrix.
- Multiply two matrices and determine the size of the product.
- Apply key properties of matrix multiplication, including distributive and associative laws, while recognizing that multiplication is not commutative.
- Express systems of linear equations in vector form and matrix form using matrix multiplication.
- Construct the augmented matrix of a system of equations from its coefficient matrix and constant vector.

In the previous section, matrix addition felt intuitive. Now we turn to the question of what it means to ***multiply*** matrices. To start, consider the product

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

What does this mean? A natural (but incorrect) guess is the entrywise product

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

However, that is ***not*** how matrix multiplication is defined. The correct product is

$$\begin{bmatrix} 5 & 3 \\ 11 & 5 \end{bmatrix}.$$

So, unlike addition, matrix multiplication does not proceed entry by entry; it follows a different rule (which we will make precise shortly). To complicate matters further—and to hint at the real rule—look at

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

Even though these matrices ***do not*** have the same size (so they cannot be added), this product ***is*** defined, and it equals

$$\begin{bmatrix} 5 & 3 & -2 \\ 11 & 5 & -4 \end{bmatrix}.$$

(Notice that multiplying a 2×2 matrix by a 2×3 matrix produced a 2×3 matrix.)

We will now step back and build the rule for **matrix multiplication** carefully. As a first step, we give an alternative definitions of the row and column matrices introduced in the previous section.

Definition: Column and Row Vectors

A row vector is a row matrix of size $1 \times n$, for example

$$a = [a_1 \quad a_2 \quad \cdots \quad a_n]. \tag{2.2.1}$$

A column vector is a column matrix of size $m \times 1$, for example

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}. \tag{2.2.2}$$

An example of a row vector is

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

and an example of a column vector is

$$b = \begin{bmatrix} 1 \\ 7 \\ 8 \end{bmatrix}.$$

Before we learn how to multiply matrices in general, we will learn what it means to multiply a row vector by a column vector.

Definition: Multiply a Row Vector by a Column Vector

Let a be a $1 \times n$ row vector with entries $a_1, a_2, \cdots, a_n$ and let b be an $n \times 1$ column vector with entries $b_1, b_2, \cdots, b_n$. The product of a and b, denoted ab, is

$$ab = [a_1 \quad a_2 \quad \cdots \quad a_n] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

One should note the following in the definition:

- The row vector a must have n columns and the column vector b **must have** n rows in order for the product ab to be defined.
- The expression $\sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$ is a scalar, or equivalently, a 1×1 matrix

Before we state the condition for multiplying matrices in general, let us first look at an example.

Example 2.2.1

Let

$$a = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, b = \begin{bmatrix} 2 & 0 & 1 & -1 \end{bmatrix}, x = \begin{bmatrix} -2 \\ 4 \\ 3 \end{bmatrix}, y = \begin{bmatrix} 1 \\ 2 \\ 5 \\ 0 \end{bmatrix}.$$

Find the following products.

$$ax, \quad by, \quad ay, \quad ab, \quad xa \tag{2.2.3}$$

Solution

$$ax = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} -2 \\ 4 \\ 3 \end{bmatrix} = 1(-2) + 2(4) + 3(3) = 15 \tag{2.2.4}$$

$$by = \begin{bmatrix} 2 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 5 \\ 0 \end{bmatrix} = 2(1) + 0(2) + 1(5) - 1(0) = 7 \tag{2.2.5}$$

ay is not defined. A row vector and column vector must have the same number of entries in order to multiply.

ab is not defined. We only know how to multiply a row vector by a column vector, not two row vectors.

xa is defined, but we have not learned how to compute it yet. At this point, we only know how to multiply a row vector by a column vector, not a column vector by a row vector. (Indeed, $ax \neq xa$!).

We notice that size is crucial when multiplying column and row matrices.

Before introducing the general definition of matrix multiplication, we say that two matrices A and B are ***multiplication conformable*** if and only if the number of columns of A equals the number of rows of B.

Definition: Matrix Multiplication

Let A be an $m \times r$ matrix, and let B be an $r \times n$ matrix. The *matrix product of A and B*, denoted $A \cdot B$, or simply AB, is the $m \times n$ matrix M whose entry in the ith row and j^{th} column is the product of the ith row of A and the jth column of B.

It may help to illustrate it in this way. Let matrix A have rows $a_1, a_2, \cdots, a_m$ and let B have columns $b_1, b_2, \cdots, b_n$. Thus A looks like

$$\begin{bmatrix} - & a_1 & - \\ - & a_2 & - \\ & \vdots & \\ - & a_m & - \end{bmatrix}$$

where the "$-$" symbols just serve as reminders that the a_i represent rows, and B looks like

$$\begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_n \\ | & | & & | \end{bmatrix}$$

where again, the "|" symbols just remind us that the b_i represent column vectors. Then

$$AB = \begin{bmatrix} a_1b_1 & a_1b_2 & \cdots & a_1b_n \\ a_2b_1 & a_2b_2 & \cdots & a_2b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_mb_1 & a_mb_2 & \cdots & a_mb_n \end{bmatrix}$$

Two quick notes about this definition. First, in order for A and B to be conformable for multiplication, the number of columns of A must be the same as the number of rows of B (these are called the "inner dimensions"). Secondly, the resulting matrix has the same number of rows as A and the same number of columns as B (these are called the "outer dimensions").

$$\overbrace{\underbrace{(m \times r)}_{\text{these inner dimensions must match}} \times (r \times n)}^{\text{final dimensions are the outer dimensions}}$$

Example 2.2.2

Revisit the matrix product we saw at the beginning of this section; multiply

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

Solution

Let's call our first matrix A and the second B. We should first check to see that we can actually perform this multiplication. Matrix A is 2×2 and B is 2×3. The "inner" dimensions match up, so we can compute the product; the "outer" dimensions tell us that the product will be 2×3. Let

$$AB = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix}.$$

Let's find the value of each of the entries.

The entry m_{11} is in the first row and first column; therefore to find its value, we need to multiply the first row of A by the first column of B. Thus

$$m_{11} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

=1(1)+2(2)=5. \nonumber \]

So now we know that

$$AB = \begin{bmatrix} 5 & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix}.$$

Finishing out the first row, we have

$$m_{12} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = 1(-1) + 2(2) = 3$$

using the first row of A and the second column of B, and

$$m_{13} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = 1(0) + 2(-1) = -2$$

using the first row of A and the third column of B. Thus we have

$$AB = \begin{bmatrix} 5 & 3 & -2 \\ m_{21} & m_{22} & m_{23} \end{bmatrix}.$$

To compute the second row of AB, we multiply with the second row of A. We find

$$m_{21} = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 11,$$

$$m_{22} = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = 5,$$

and

$$m_{23} = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = -4.$$

Thus

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

Example 2.2.3

Multiply

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

Solution

Let's first check to make sure this product is defined. Again calling the first matrix A and the second B, we see that A is a 3×2 matrix and B is a 2×4 matrix; the inner dimensions match so the product is defined, and the product will be a 3×4 matrix,

$$AB = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix}.$$

We will demonstrate how to compute some of the entries, then give the final answer. The reader can fill in the details of how each entry was computed.

$$m_{11} = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = -1.$$

$$m_{13} = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 7 \end{bmatrix} = -6.$$

$$m_{23} = \begin{bmatrix} 5 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 7 \end{bmatrix} = 19.$$

$$m_{24} = \begin{bmatrix} 5 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 9 \end{bmatrix} = 23.$$

$$m_{32} = \begin{bmatrix} -2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \end{bmatrix} = 16.$$

$$m_{34} = \begin{bmatrix} -2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 9 \end{bmatrix} = 25.$$

So far, we've computed this much of AB:

$$AB = \begin{bmatrix} -1 & m_{12} & -6 & m_{14} \\ m_{21} & m_{22} & 19 & 23 \\ m_{31} & 16 & m_{33} & 25 \end{bmatrix}.$$

The final product is

$$AB = \begin{bmatrix} -1 & -5 & -6 & -8 \\ 9 & 17 & 19 & 23 \\ 4 & 16 & 19 & 25 \end{bmatrix}.$$

Example 2.2.4

Multiply, if possible

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

Solution

Notice that \(A)\ and \(B)\ are not conformable.

Example 2.2.5

In Example 2.2.1, we were told that the product $\vec{x}\vec{u}$ was defined, where

$$\vec{x} = \begin{bmatrix} -2 \\ 4 \\ 3 \end{bmatrix} \quad \text{and} \quad \vec{u} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix},$$

although we were not shown what the product was. Find $\vec{x}\vec{u}$.

Solution

Again, we need to check to make sure the dimensions work correctly (remember that even though we are referring to u and x as vectors, they are, in fact, just matrices).

The column vector x has dimensions 3×1, whereas the row vector u has dimensions 1×3. Since the inner dimensions do match, the matrix product is defined; the outer dimensions tell us that the product will be a 3×3 matrix, as shown below:

$$xu = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}.$$

To compute the entry m_{11}, we multiply the first row of x by the first column of u. What is the first row of x? It is just the number -2. What is the first column of u? It is the number 1. Thus $m_{11} = -2$.

For the entry m_{12}, we multiply the first row of x by the second column of u; that is, we compute $-2(2)$. So $m_{12} = -4$.

For the entry m_{23}, we multiply the second row of x by the third column of u; that is, $4(3)$, so $m_{23} = 12$.

One final example: m_{31} comes from multiplying the third row of x, which is 3, by the first column of u, which is 1. Therefore $m_{31} = 3$.

So far we have computed

$$xu = \begin{bmatrix} -2 & -4 & m_{13} \\ m_{21} & m_{22} & 12 \\ 3 & m_{32} & m_{33} \end{bmatrix}.$$

After performing all 9 multiplications, we find

$$xu = \begin{bmatrix} -2 & -4 & -6 \\ 4 & 8 & 12 \\ 3 & 6 & 9 \end{bmatrix}.$$

In this last example, we saw a "nonstandard" multiplication (at least, it felt nonstandard). Studying the entries of this matrix, it seems that there are several different patterns that can be seen amongst the entries. (Remember that mathematicians like to look for patterns. Also remember that we often guess wrong at first; don't be scared and try to identify some patterns.)

In Section 2.1, we identified the zero matrix $\mathbf{0}$ that had a nice property in relation to matrix addition (i.e., $A + \mathbf{0} = A$ for any matrix A). In the following example we'll identify a matrix that works well with multiplication as well as some multiplicative properties. For instance, we've learned how $1 \cdot A = A$; is there a *matrix* that acts like the number 1? That is, can we find a matrix X where $X \cdot A = A$?[3]

Example 2.2.6

Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$
$$C = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}, \quad I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Find the following products.

1. AB
2. BA
3. $A\mathbf{0}_{3\times 4}$
4. AI
5. IA
6. I^2
7. BC
8. B^2

Solution

We will find each product, but we leave the details of each computation to the reader.

1. $AB = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 6 & 6 \\ 0 & 0 & 0 \\ -7 & -7 & -7 \end{bmatrix}$
2. $BA = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix} = \begin{bmatrix} 1 & -13 & 11 \\ 1 & -13 & 11 \\ 1 & -13 & 11 \end{bmatrix}$
3. $A\mathbf{0}_{3\times 4} = \mathbf{0}_{3\times 4}$.
4. $AI = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix}$
5. $IA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix}$
6. We haven't formally defined what I^2 means, but we could probably make the reasonable guess that $I^2 = I \cdot I$. Thus

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

7. $BC = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \\ 3 & 3 & 3 \end{bmatrix}$
8. $B^2 = BB = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \\ 3 & 3 & 3 \end{bmatrix}$

This example is simply chock full of interesting ideas; it is almost hard to think about where to start.

Interesting Idea #1: Notice that in our example, $AB \neq BA$! When dealing with numbers, we were used to the idea that $ab = ba$. With matrices, multiplication is *not* commutative. (Of course, we can find special situations where it does work. In general, though, it doesn't.)

Interesting Idea #2: Right before this example we wondered if there was a matrix that "acted like the number 1," and guessed it may be a matrix of all 1s. However, we found out that such a matrix does not work in that way; in our example, $AB \neq A$. We did find that $AI = IA = A$. There is a Multiplicative Identity; it just isn't what we thought it would be. And just as $1^2 = 1$, $I^2 = I$.

Interesting Idea #3: When dealing with numbers, we are very familiar with the notion that "If $ax = bx$, then $a = b$." (As long as $x \neq 0$.) Notice that, in our example, $BB = BC$, yet $B \neq C$. In general, just because $AX = BX$, we *cannot* conclude that $A = B$. Matrix multiplication is turning out to be a very strange operation. We are very used to multiplying numbers, and we know a bunch of properties that hold when using this type of multiplication. When multiplying matrices, though, we probably find ourselves asking two questions, "What *does* work?" and "What *doesn't* work?" We'll answer these questions; first we'll do an example that demonstrates some of the things that do work.

Example 2.2.7

Let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Find the following:

1. $A(B + C)$
2. $AB + AC$
3. $A(BC)$
4. $(AB)C$

Solution

We'll compute each of these without showing all the intermediate steps. Keep in mind order of operations: things that appear inside of parentheses are computed first.

1.
$$\begin{aligned} A(B+C) &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \right) \\ &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 7 & 4 \\ 17 & 10 \end{bmatrix} \end{aligned}$$

2.
$$\begin{aligned} AB + AC &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 3 & -1 \\ 7 & -1 \end{bmatrix} + \begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix} \\ &= \begin{bmatrix} 7 & 4 \\ 17 & 10 \end{bmatrix} \end{aligned}$$

3.
$$\begin{aligned} A(BC) &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right) \\ &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 3 & 3 \\ 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 1 \\ 13 & 5 \end{bmatrix} \end{aligned}$$

4.
$$\begin{aligned} (AB)C &= \left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\right)\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 3 & -1 \\ 7 & -1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 1 \\ 13 & 5 \end{bmatrix} \end{aligned}$$

Looking back at our example, two important patterns appear. First, the distributive property seems to hold; that is, $A(B+C) = AB + AC$. This is encouraging, since many of the algebraic rules we are used to from ordinary algebra still apply to matrices. Second, the associative property also seems to hold; that is, $A(BC) = (AB)C$. This means that when multiplying several matrices, we do not have to worry about the order in which we carry out the pairwise products.

With these observations in place, we are ready to move toward an important theorem. To do so, let us recall a matrix we encountered in an earlier example.

In the previous section we introduced an important type of matrix. We now give its official definition.

Definition: Identity Matrix

The $n \times n$ matrix with 1's on the diagonal and zeros elsewhere is the $n \times n$ *identity matrix*, denoted I_n. When the context makes the dimension of the identity clear, the subscript is generally omitted.

Just as the zero matrix can take many different shapes and sizes, the identity matrix has a special property: it is always square. A few examples are shown below.

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Up to this point, we have experimented with examples to see what works and what does not. Examples are helpful, but they have limits. If someone claimed that $AB = BA$ always holds, it would only take one counterexample to show that the claim is false. On the other hand, if someone claimed that $A(B+C) = AB + AC$ is always true, one example is not enough to prove the claim; we need something stronger—a proof.

In this text we will not provide full proofs for every statement. The important point to remember is that whenever we present a theorem, there is a proof behind it. Our confidence comes from more than just trying out examples.

With that in mind, let us now turn to the good news: the properties of matrix multiplication that always work.

Theorem 2.2.1

Properties of Matrix Multiplication

Let A, B and C be matrices with dimensions so that the following operations make sense, and let k be a scalar. The following equalities hold:

1. $A(BC) = (AB)C$ (Associative Property)
2. $A(B + C) = AB + AB$ and
   $(B + C)A = BA + CA$ (Distributive Property)
3. $k(AB) = (kA)B = A(kB)$
4. $AI = IA = A$

The box above contained some very good news—and perhaps some surprising news. Matrix multiplication may seem like a strange operation, so it would not be shocking to learn that $A(BC) \neq (AB)C$. Fortunately, the associative property does hold, and this makes our work with matrices much easier.

Before moving on, we raise one more issue of notation. We define $A^0 = I$. If n is a positive integer, we define

$$A^n = \underbrace{A \cdot A \cdot \cdots \cdot A}_{n \text{ times}}.$$

When working with numbers, we are used to the fact that $a^{-n} = \frac{1}{a^n}$. Do negative exponents make sense with matrices as well? The answer is yes, but with some care. We will study this more fully after defining the inverse of a matrix. For now, it is important to remember that $A^{-1} \neq \frac{1}{A}$, since division of matrices does not make sense in the same way it does for numbers.

There are also a few things that do not work with matrix multiplication. The good news is that the list is short:

- Matrix multiplication is not commutative; in general, $AB \neq BA$.
- Even if $AX = BX$, we cannot conclude that $A = B$.

These facts can lead to pitfalls. For example, we are familiar with

$$(a + b)^2 = a^2 + 2ab + b^2.$$

But what about $(A + B)^2$? Here,

$$(A + B)^2 \neq A^2 + 2AB + B^2.$$

With these properties in mind, we are now prepared to revisit systems of linear equations and view them through the lens of matrices.

Systems of Linear Equations Matrix multiplication

Thinking of matrix multiplication gives us a new perspective on systems of linear equations. Instead of writing each equation separately, we can represent an entire system more compactly using vectors and matrices. This approach not only shortens notation but also reveals deeper connections between linear equations and the structure of matrix algebra. In this section, we will see how to express systems of equations in both vector form and matrix form.

\textbf{Vector Form of a System of Linear Equations}

Definition:Vector Form of a System of Linear Equation

Suppose we have a system of equations

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

This system can be written in vector form as

$$x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}. \tag{2.2.6}$$

Notice that each column vector here comes directly from the corresponding coefficient matrix. There is one column vector for each variable, along with the constant vector on the right-hand side.

This observation leads us naturally to the first important case of matrix multiplication: multiplying a matrix by a column vector. For example, consider

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}. \tag{2.2.7}$$

This equals

$$7 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + 8 \begin{bmatrix} 2 \\ 5 \end{bmatrix} + 9 \begin{bmatrix} 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 50 \\ 122 \end{bmatrix}. \tag{2.2.8}$$

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** is $AX = B$

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

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.

These forms will be very useful later when we introduce the inverse of a matrix, vector spaces, and linear transformations. They

give us a compact way to represent systems and highlight the structure behind the equations. For now, let us look at an example to see how this works in practice.

Example 2.2.1

Given $A = \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 2 & 1 & -2 \\ 2 & 1 & 4 & 1 \end{bmatrix}$ as the coefficient matrix and $X = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$.

Solution

To write the system of linear equation, we need the column vector for the constant, $b). Thus first we compute the right-hand side \(b = AX$:

$$b = \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 2 & 1 & -2 \\ 2 & 1 & 4 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1(1) + 2(2) + 1(0) + 3(1) \\ 0(1) + 2(2) + 1(0) - 2(1) \\ 2(1) + 1(2) + 4(0) + 1(1) \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \\ 5 \end{bmatrix}. \tag{2.2.9}$$

The augmented matrix $[A \mid b]$ for the system $AX = b$ is

$$[A \mid b] = \left[\begin{array}{cccc|c} 1 & 2 & 1 & 3 & 8 \\ 0 & 2 & 1 & -2 & 2 \\ 2 & 1 & 4 & 1 & 5 \end{array}\right]. \tag{2.2.10}$$

Exercise

In Exercises 2.2.1 - 2.2.12, row and column vectors $\vec{u}$ and $\vec{v}$ are defined. Find the product $\vec{u}\vec{v}$, where possible.

Exercise 2.2.1

$u = \begin{bmatrix} 1 & -4 \end{bmatrix} \quad v = \begin{bmatrix} -2 \\ 5 \end{bmatrix}$

Answer available at the end of this section.

Exercise 2.2.2

$u = \begin{bmatrix} 2 & 3 \end{bmatrix} \quad v = \begin{bmatrix} 7 \\ -4 \end{bmatrix}$

Answer available at the end of this section.

Exercise 2.2.3

$\vec{u} = \begin{bmatrix} 1 & -1 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$

Answer available at the end of this section.

Exercise 2.2.4

$a = \begin{bmatrix} 0.6 & 0.8 \end{bmatrix} \quad b = \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix}$

Answer available at the end of this section.

Exercise 2.2.5

$b = \begin{bmatrix} 1 & 2 & -1 \end{bmatrix} \quad d = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$

Answer available at the end of this section.

Exercise 2.2.6

$a = \begin{bmatrix} 3 & 2 & -2 \end{bmatrix} \quad b = \begin{bmatrix} -1 \\ 0 \\ 9 \end{bmatrix}$

Answer available at the end of this section.

Exercise 2.2.7

$a = \begin{bmatrix} 8 & -4 & 3 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$

Answer available at the end of this section.

Exercise 2.2.8

$a = \begin{bmatrix} -3 & 6 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

Answer available at the end of this section.

Exercise 2.2.9

$a = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$

Answer available at the end of this section.

Exercise 2.2.10

$$b = \begin{bmatrix} 6 & 2 & -1 & 2 \end{bmatrix} \quad c = \begin{bmatrix} 3 \\ 2 \\ 9 \\ 5 \end{bmatrix}$$

Answer available at the end of this section.

Answer available at the end of this section.

Exercise 2.2.11

$$a = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \quad b = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

Answer available at the end of this section.

Answer available at the end of this section.

Exercise 2.2.12

$$\vec{u} = \begin{bmatrix} 2 & -5 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Answer available at the end of this section.

Answer available at the end of this section.

In Exercises 2.2.13 - 2.2.27, matrices A and B are defined.

a. Give the dimensions of A and B. If the dimensions properly match, give the dimensions of AB and BA.
b. Find the products AB and BA, if possible.

Exercise 2.2.13

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

Answer available at the end of this section.

Exercise 2.2.14

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

Answer available at the end of this section.

Exercise 2.2.15

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

Answer available at the end of this section.

Exercise 2.2.16

$$A = \begin{bmatrix} 0 & 1 \\ 1 & -1 \\ -2 & -4 \end{bmatrix} \quad B = \begin{bmatrix} -2 & 0 \\ 3 & 8 \end{bmatrix}$$

Answer available at the end of this section.

Exercise 2.2.17

$$A = \begin{bmatrix} 9 & 4 & 3 \\ 9 & -5 & 9 \end{bmatrix} \quad B = \begin{bmatrix} -2 & 5 \\ -2 & -1 \end{bmatrix}$$

Answer available at the end of this section.

Exercise 2.2.18

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

Answer available at the end of this section.

Exercise 2.2.19

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

Answer available at the end of this section.

Exercise 2.2.20

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

Answer available at the end of this section.

Exercise 2.2.21

$$A = \begin{bmatrix} 8 & -2 \\ 4 & 5 \\ 2 & -5 \end{bmatrix} \quad B = \begin{bmatrix} -5 & 1 & -5 \\ 8 & 3 & -2 \end{bmatrix}$$

Answer available at the end of this section.

Exercise 2.2.22

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

Answer available at the end of this section.

Exercise 2.2.23

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

Answer available at the end of this section.

Exercise 2.2.24

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

Answer available at the end of this section.

Exercise 2.2.25

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

Answer available at the end of this section.

Exercise 2.2.26

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

Answer available at the end of this section.

Exercise 2.2.27

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

Answer available at the end of this section.

In Exercises 2.2.28 - 2.2.33, a *diagonal* matrix D and a matrix A are given. Find the products DA and AD, where possible.

Exercise 2.2.28

$$D = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} \quad A = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix}$$

Answer available at the end of this section.

Exercise 2.2.29

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

Answer available at the end of this section.

Exercise 2.2.30

$$D = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Answer available at the end of this section.

Exercise 2.2.31

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

Answer available at the end of this section.

Exercise 2.2.32

$$D = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} \quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Answer available at the end of this section.

Exercise 2.2.33

$$D = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \quad A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

Answer available at the end of this section.

In Exercises 2.2.34 - 2.2.39, a matrix A and a vector $\vec{x}$ are given. Find the product $A\vec{x}$.

Exercise 2.2.34

$$A = \begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix}, \vec{x} = \begin{bmatrix} 4 \\ 9 \end{bmatrix}$$

Answer available at the end of this section.

Exercise 2.2.35

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

Answer available at the end of this section.

Exercise 2.2.36

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

Answer available at the end of this section.

Exercise 2.2.37

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

Answer available at the end of this section.

Exercise 2.2.38

$$A = \begin{bmatrix} 2 & -1 \\ 4 & 3 \end{bmatrix}, \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Answer available at the end of this section.

Exercise 2.2.39

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 2 \\ 2 & 3 & 1 \end{bmatrix}, \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Answer available at the end of this section.

Exercise 2.2.40

Let $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Find A^2 and A^3.

Answer available at the end of this section.

Exercise 2.2.41

Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$. Find A^2 and A^3.

Answer available at the end of this section.

Exercise 2.2.42

Let $A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$. Find A^2 and A^3.

Answer available at the end of this section.

Exercise 2.2.43

Let $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$. Find A^2 and A^3.

Answer available at the end of this section.

Exercise 2.2.44

Let $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. Find A^2 and A^3.

Answer available at the end of this section.

Exercise 2.2.45

In the text, we state that $(A + B)^2 \neq A^2 + 2AB + B^2$. We investigate that claim here.

a. Let $A = \begin{bmatrix} 5 & 3 \\ -3 & -2 \end{bmatrix}$ and let $B = \begin{bmatrix} -5 & -5 \\ -2 & 1 \end{bmatrix}$. Compute $A + B$.
b. Find $(A + B)^2$ by using your answer from (a).
c. Compute $A^2 + 2AB + B^2$.
d. Are the results from (a) and (b) the same?
e. Carefully expand the expression $(A + B)^2 = (A + B)(A + B)$ and show why this is not equal to $A^2 + 2AB + B^2$.

Answer available at the end of this section.

Answers

Exercise 2.2.1

-22

Exercise 2.2.2

2

Exercise 2.2.3

0

Exercise 2.2.4

1

Exercise 2.2.5

5

Exercise 2.2.6

-21

Exercise 2.2.7

15

Exercise 2.2.8

-8

Exercise 2.2.9

-2

Exercise 2.2.10

Exercise 2.2.10

23

Exercise 2.2.11

Exercise 2.2.11

Not possible.

Exercise 2.2.12

Exercise 2.2.12

Not possible.

Exercise 2.2.13

$AB = \begin{bmatrix} 8 & 3 \\ 10 & -9 \end{bmatrix}$

$BA = \begin{bmatrix} -3 & 24 \\ 4 & 2 \end{bmatrix}$

Exercise 2.2.14

$AB = \begin{bmatrix} 24 & -24 \\ 17 & -17 \end{bmatrix}$

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

Exercise 2.2.15

$AB = \begin{bmatrix} -1 & -2 & 12 \\ 10 & 4 & 32 \end{bmatrix}$

BA is not possible.

Exercise 2.2.16

$$AB = \begin{bmatrix} 3 & 8 \\ -5 & -8 \\ -8 & -32 \end{bmatrix}$$

BA is not possible.

Exercise 2.2.17

AB is not possible.

$$BA = \begin{bmatrix} 27 & -33 & 39 \\ -27 & -3 & -15 \end{bmatrix}$$

Exercise 2.2.18

$$AB = \begin{bmatrix} 10 & -18 & 11 \\ -45 & 24 & -21 \\ -15 & 12 & -9 \end{bmatrix}$$

$$BA = \begin{bmatrix} 52 & -21 \\ 45 & -27 \end{bmatrix}$$

Exercise 2.2.19

$$AB = \begin{bmatrix} -32 & 34 & -24 \\ -32 & 38 & -8 \\ -16 & 21 & 4 \end{bmatrix}$$

$$BA = \begin{bmatrix} 22 & -14 \\ -4 & -12 \end{bmatrix}$$

Exercise 2.2.20

$$AB = \begin{bmatrix} -10 & 19 & -32 \\ 10 & 31 & -28 \\ 20 & 37 & -26 \end{bmatrix}$$

$$BA = \begin{bmatrix} -5 & -14 \\ 45 & 0 \end{bmatrix}$$

Exercise 2.2.21

$$AB = \begin{bmatrix} -56 & 2 & -36 \\ 20 & 19 & -30 \\ -50 & -13 & 0 \end{bmatrix}$$

$$BA = \begin{bmatrix} -46 & 40 \\ 72 & 9 \end{bmatrix}$$

Exercise 2.2.22

$$AB = \begin{bmatrix} -7 & 3 & 7 & -15 \\ -5 & -1 & -17 & 5 \end{bmatrix}$$

BA is not possible.

Exercise 2.2.23

$$AB = \begin{bmatrix} -15 & -22 & -21 & -1 \\ 16 & -53 & -59 & -31 \end{bmatrix}$$

BA is not possible.

Exercise 2.2.24

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

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

Exercise 2.2.25

$$AB = \begin{bmatrix} 0 & 0 & 4 \\ 6 & 4 & -2 \\ 2 & -4 & -6 \end{bmatrix}$$

$$BA = \begin{bmatrix} 2 & -2 & 6 \\ 2 & 2 & 4 \\ 4 & 0 & -6 \end{bmatrix}$$

Exercise 2.2.26

$$AB = \begin{bmatrix} 0 & -44 & 18 \\ -20 & -1 & -18 \\ -5 & -21 & -3 \end{bmatrix}$$

$$BA = \begin{bmatrix} -25 & -5 & -25 \\ 25 & -11 & 2 \\ -15 & 19 & 32 \end{bmatrix}$$

Exercise 2.2.27

$$AB = \begin{bmatrix} 21 & -17 & -5 \\ 19 & 5 & 19 \\ 5 & 9 & 4 \end{bmatrix}$$

$$BA = \begin{bmatrix} 19 & 5 & 23 \\ 5 & -7 & -1 \\ -14 & 6 & 18 \end{bmatrix}$$

Exercise 2.2.28

$$DA = \begin{bmatrix} 6 & -4 \\ 18 & -8 \end{bmatrix}$$

$$AD = \begin{bmatrix} 6 & 12 \\ -6 & -8 \end{bmatrix}$$

Exercise 2.2.29

$$DA = \begin{bmatrix} 4 & -6 \\ 4 & -6 \end{bmatrix}$$

$$AD = \begin{bmatrix} 4 & 8 \\ -3 & -6 \end{bmatrix}$$

Exercise 2.2.30

$$DA = \begin{bmatrix} -1 & 4 & 9 \\ -4 & 10 & 18 \\ -7 & 16 & 27 \end{bmatrix}$$

$$AD = \begin{bmatrix} -1 & -2 & -3 \\ 8 & 10 & 12 \\ 21 & 24 & 27 \end{bmatrix}$$

Exercise 2.2.31

$$DA = \begin{bmatrix} 2 & 2 & 2 \\ -6 & -6 & -6 \\ -15 & -15 & -15 \end{bmatrix}$$

$$AD = \begin{bmatrix} 2 & -3 & 5 \\ 4 & -6 & 10 \\ -6 & 9 & -15 \end{bmatrix}$$

Exercise 2.2.32

$$DA = \begin{bmatrix} d_1a & d_1b \\ d_2c & d_2d \end{bmatrix}$$

$$AD = \begin{bmatrix} d_1a & d_2b \\ d_1c & d_2d \end{bmatrix}$$

Exercise 2.2.33

$$DA = \begin{bmatrix} d_1a & d_1b & d_1c \\ d_2d & d_2e & d_2f \\ d_3g & d_3h & d_3i \end{bmatrix}$$

$$AD = \begin{bmatrix} d_1a & d_2b & d_3c \\ d_1d & d_2e & d_3f \\ d_1g & d_2h & d_3i \end{bmatrix}$$

Exercise 2.2.34

$$A\vec{x} = \begin{bmatrix} 35 \\ -5 \end{bmatrix}$$

Exercise 2.2.35

$$A\vec{x} = \begin{bmatrix} -6 \\ 11 \end{bmatrix}$$

Exercise 2.2.36

$$A\vec{x} = \begin{bmatrix} 8 \\ 7 \\ 3 \end{bmatrix}$$

Exercise 2.2.37

$$A\vec{x} = \begin{bmatrix} -5 \\ 5 \\ 21 \end{bmatrix}$$

Exercise 2.2.38

$$A\vec{x} = \begin{bmatrix} 2x_1 - x_2 \\ 4x_1 + 3x_2 \end{bmatrix}$$

Exercise 2.2.39

$$A\vec{x} \begin{bmatrix} x_1 + 2x_2 + 3x_3 \\ x_1 + 2x_3 \\ 2x_1 + 3x_2 + x_3 \end{bmatrix}$$

Exercise 2.2.40

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

Exercise 2.2.41

$A^2 = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}; A^3 = \begin{bmatrix} 8 & 0 \\ 0 & 27 \end{bmatrix}$

Exercise 2.2.42

$A^2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 25 \end{bmatrix}; A^3 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 27 & 0 \\ 0 & 0 & 125 \end{bmatrix}$

Exercise 2.2.43

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

Exercise 2.2.44

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

Exercise 2.2.45

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

b. $\begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix}$

c. $\begin{bmatrix} -11 & -15 \\ 37 & 32 \end{bmatrix}$

d. No

e. $(A+B)(A+B) = AA + AB + BA + BB = A^2 + AB + BA + B^2$

This page titled 2.2: Matrix Multiplication is shared under a license and was authored, remixed, and/or curated by LibreTexts.

- 2.2: Matrix Multiplication by Gregory Hartman et al. is licensed CC BY-NC 3.0. Original source: https://github.com/APEXCalculus/Fundamentals-of-Matrix-Algebra.
- 2.2: Multiplication of Matrices by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.
- 2.2.1: Exercises 2.2 by Gregory Hartman et al. is licensed CC BY-NC 3.0.

2.3: 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 2.3.1: 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 available at the end of this section.

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

Definition 2.3.1: 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 2.3.1: 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 available at the end of this section.

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

Example 2.3.1: 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 2.3.2: 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

$$\begin{aligned} \begin{bmatrix} 0 \\ 0 \end{bmatrix} &= A^{-1}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) \\ &= A^{-1}\left(A\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) \\ &= \left(A^{-1}A\right)\begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= I\begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ 1 \end{bmatrix} \end{aligned} \tag{2.3.1}$$

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.

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 2.3.1: 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 available at the end of this section.

We now consider a second important lemma.

Lemma 2.3.2: 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 available at the end of this section.

An important theorem follows from this lemma.

Theorem 2.3.1: Invertible Matrices are Square

Only square matrices can be invertible.

Proof available at the end of this section.

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 2.3.1: 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 2.3.2: 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 available at the end of this section.

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 2.3.1: 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

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2.3.2}$$

Therefore, $A^T A = I_2$, where I_2 is the 2×2 identity matrix. However, the product AA^T is

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2.3.3}$$

Hence AA^T is not the 3×3 identity matrix. This shows that for Theorem 2.3.2, 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 2.3.3: 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 available at the end of this section.

Theorem 2.3.3 corresponds to Algorithm 2.7.1, which claims that A^{-1} is found by row reducing the augmented matrix $[A|I]$ to the form $[I|A^{-1}]$. This will be a matrix product $E[A|I]$ where E is a product of elementary matrices. By the rules of matrix multiplication, we have that $E[A|I] = [EA|EI] = [EA|E]$.

It follows that the reduced row-echelon form of $[A|I]$ is $[EA|E]$, where EA gives the reduced row-echelon form of A. By Theorem 2.3.3, if $EA \neq I$, then A is not invertible, and if $EA = I$, A is invertible. If $EA = I$, then by Theorem 2.3.2, $E = A^{-1}$. This proves that Algorithm 2.7.1 does in fact find A^{-1}.

Answers

Lemma 2.3.1: Multiplication by the Identity Matrix

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.

Theorem 2.3.1: Uniqueness of Inverse

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.

Lemma 2.3.1: Invertible Matrix and Zeros

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.

Lemma 2.3.2: Size of Invertible Matrix

Let R be the reduced row-echelon form of A. By Lemma 2.3.1, 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.

Theorem 2.3.1: Invertible Matrices are Square

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 2.3.2 with A and B, to obtain the inequality $m \leq n$. We then apply Lemma 2.3.2 again (switching the order of the matrices), to obtain the inequality $n \leq m$. It follows that $m = n$, as we wanted.

Theorem 2.3.2: Unique Inverse of a Matrix

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 2.3.1 gives us that R does not have a row of zeros. By noting that R is a square matrix and applying Proposition 2.3.1, 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

$$\begin{aligned} BA = IBIA &= (EA)B(E^{-1}E)A \\ &= E(AB)E^{-1}(EA) \\ &= EIE^{-1}I \\ &= EE^{-1} = I \end{aligned} \tag{2.3.4}$$

It follows from the definition of the inverse of a matrix that $B = A^{-1}$ and $A = B^{-1}$.

Theorem 2.3.3: The reduced row-echelon form of an Invertible Matrix

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 2.3.1, we get that the of A does not have a row of zeros. Then by Theorem 2.3.1, it follows that A and the reduced row-echelon form of A are square matrices. Finally, by Proposition 2.3.1, 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 2.3.2, we can conclude that A is invertible.

This page titled 2.3: 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.
- 2.9: More on Matrix Inverses by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

2.4: 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

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

for the first system and

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{2.4.1}$$

for the second.

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

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

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

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

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

Now solve the second system, (2.4.1) 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

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

and row reduced until we obtained

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

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

This exploration motivates the following important algorithm.

Algorithm 2.4.1: 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 2.4.1: 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

$$[A|I] = \begin{bmatrix} 1 & 2 & 2 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 \\ 3 & 1 & -1 & 0 & 0 & 1 \end{bmatrix}$$

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

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

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

$$\begin{bmatrix} 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & -10 & 0 & -5 & 5 & 0 \\ 0 & 0 & 14 & 1 & 5 & -2 \end{bmatrix}$$

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

$$\begin{bmatrix} -7 & -14 & 0 & -6 & 5 & -2 \\ 0 & -10 & 0 & -5 & 5 & 0 \\ 0 & 0 & 14 & 1 & 5 & -2 \end{bmatrix}$$

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

$$\begin{bmatrix} -7 & 0 & 0 & 1 & -2 & -2 \\ 0 & -10 & 0 & -5 & 5 & 0 \\ 0 & 0 & 14 & 1 & 5 & -2 \end{bmatrix}$$

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

$$\begin{bmatrix} 1 & 0 & 0 & -\frac{1}{7} & \frac{2}{7} & \frac{2}{7} \\ 0 & 1 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 & \frac{1}{14} & \frac{5}{14} & -\frac{1}{7} \end{bmatrix}$$

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

$$\begin{bmatrix} -\frac{1}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{14} & \frac{5}{14} & -\frac{1}{7} \end{bmatrix}$$

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 2.4.2: 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 $[A|I]$

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

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

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

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

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

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 2.4.3: 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

$$AX = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = B \tag{2.4.2}$$

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 (2.4.2) 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 (2.4.2) 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 2.4.1: Inverses of Transposes and Products

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

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

Consider the following theorem.

Theorem 2.4.2: 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 2.4: 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.

2.5: 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 2.5.1: 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 2.5.1: 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 2.5.1: 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 2.5.1: 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 2.5.1 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 2.5.2: 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 2.5.2: 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 2.5.3: 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 2.5.3: 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 2.5.2: 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 2.5.4: 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 2.5.2: 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 2.5.5: 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

$$\begin{aligned} B &= E(-2 \times 1 + 2)\left(P^{12}A\right) \\ &= \left(E(-2 \times 1 + 2)P^{12}\right)A \\ &= UA \end{aligned} \tag{2.5.1}$$

It remains to find the matrix U.

$$\begin{aligned} U &= E(-2 \times 1 + 2)P^{12} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & -2 & 1 \end{bmatrix} \end{aligned} \tag{2.5.2}$$

We can verify that $B = UA$ holds for this matrix U:

$$\begin{aligned} UA &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \\ &= B \end{aligned} \tag{2.5.3}$$

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 2.5.3: 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 2.5.3.

Example 2.5.6: The Form $B = UA$, Revisited

Let $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 2 & 0 \end{bmatrix}$. Using the process outlined in Theorem 2.5.3, find U such that $B = UA$.

Solution

First, set up the matrix $[A|I_m]$.

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

Now, row reduce this matrix until the left side equals the reduced row-echelon form of A.

$$\begin{aligned} \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & 0 & 1 \end{bmatrix} &\rightarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 \end{bmatrix} \\ &\rightarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -2 & 1 \end{bmatrix} \end{aligned} \tag{2.5.4}$$

The left side of this matrix is B, and the right side is U. Comparing this to the matrix U found above in Example 2.5.5, 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.

$$\begin{aligned} B &= UA \\ I_n &= UA \\ U^{-1} &= A \end{aligned} \tag{2.5.5}$$

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

$$\begin{aligned} A &= U^{-1} \\ &= E_k^{-1} \cdots E_2^{-1} E_1^{-1} \end{aligned} \tag{2.5.6}$$

and A can be written as a product of elementary matrices.

Theorem 2.5.4: 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 2.5.7: 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 2.5.3 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:

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

represented by the elementary matrix $E_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Secondly:

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

represented by the elementary matrix $E_2 = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Finally:

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

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.

$$\begin{aligned} U^{-1} &= (E_3E_2E_1)^{-1} \\ &= E_1^{-1}E_2^{-1}E_3^{-1} \\ &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \\ &= A \end{aligned} \tag{2.5.7}$$

This gives A written as a product of elementary matrices. By Theorem 2.5.4 it follows that A is invertible.

This page titled 2.5: 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.

2.6: 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 2.6.1: 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 2.6.2: 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 2.6.3: 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 2.6.4: 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

$$X = \begin{bmatrix} -\frac{3}{5} + \frac{7}{5}t \\ \frac{9}{5} - \frac{11}{5}t \\ t \\ -1 \end{bmatrix}, \ t \in \mathbb{R}.$$

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 2.6.1: 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

$$L = \begin{bmatrix} 1 & & \\ a_1 & 1 & \\ \vdots & & \ddots \\ a_n & & & 1 \end{bmatrix} \tag{2.6.1}$$

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 available at the end of this section.

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 2.6.1, 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.

Answers

Lemma 2.6.1: Multiplier Method and Triangular Matrices

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 (2.6.1) or the single nonzero column is in another position, multiplication by -1 as described in the lemma clearly results in L^{-1}.

This page titled 2.6: 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.

3: DETERMINANTS

3.1: Basic Techniques

3.1E: Exercises for Section 3.1

3.2: Properties of Determinants

3.2E: Exercises for Section 3.2

3.3: Application of the Determinant to Inverses; Cramer's Rule

3.3E: Exercises for Section 3.3

3.4: Determinants and Geometry

3.4E: Exercises for Section 3.4

This page titled 3: Determinants is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, De Anza College.

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

3.1: Basic Techniques

Learning Objectives

- Evaluate the determinant of a square matrix using Laplace Expansion (cofactor expansion)

Cofactors and 2 x 2 Determinants

Let A be an $n \times n$ matrix. That is, let A be a square matrix. The **determinant** of A, denoted by $\det(A)$ is a very important number which we will explore throughout this section.

If A is a 2×2 matrix, the determinant is given by the following formula.

Definition 3.1.1: 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 $\text{minor}(A)_{ij}$, 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 $\text{minor}\,(A)_{12}$ and $\text{minor}\,(A)_{23}$.

Solution

First we will find $\text{minor}\,(A)_{12}$. 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

$$\text{minor}\,(A)_{12} = \det \begin{bmatrix} 4 & 2 \\ 3 & 1 \end{bmatrix}$$

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 $\text{minor}\,(A)_{12} = -2$.

Similarly, $\text{minor}\,(A)_{23}$ is the determinant of the 2×2 matrix which results when you delete the second row and the third column. This minor is therefore

$$\text{minor}\,(A)_{23} = \det \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} = -4$$

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 $\text{cof}\,(A)_{ij}$ is defined to be

$$\text{cof}\,(A)_{ij} = (-1)^{i+j} \text{minor}\,(A)_{ij}$$

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 $\text{cof}\,(A)_{ij}$.

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 $\text{cof}\,(A)_{12}$ and $\text{cof}\,(A)_{23}$.

Solution

We will use Definition 3.1.3 to compute these cofactors.

First, we will compute $\text{cof}\,(A)_{12}$. Therefore, we need to find $\text{minor}\,(A)_{12}$. This is the determinant of the 2×2 matrix which results when you delete the first row and the second column. Thus $\text{minor}\,(A)_{12}$ is given by

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

Then,

$$\text{cof}\,(A)_{12} = (-1)^{1+2}\,\text{minor}\,(A)_{12} = (-1)^{1+2}\,(-2) = 2$$

Hence, $\text{cof}\,(A)_{12} = 2$.

Similarly, we can find $\text{cof}\,(A)_{23}$. First, find $\text{minor}\,(A)_{23}$, 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,

$$\text{cof}\,(A)_{23} = (-1)^{2+3}\,\text{minor}\,(A)_{23} = (-1)^{2+3}\,(-4) = 4$$

You may wish to find the remaining cofactors for the above matrix. Remember that there is a cofactor for every entry in the matrix.

Watch the accompanying video for a visual demonstration of the this concept.

- **Video Length**: 6 minutes 44 seconds
- **Context**: This video demonstrates how to find the minors and cofactors of a 3×3 matrix: $\begin{bmatrix} 1 & 3 & 2 \\ -2 & 4 & 5 \\ 0 & -6 & 9 \end{bmatrix}$

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

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.

$$\det(A) = \overbrace{1(-1)^{1+1}\begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{11}} + \overbrace{4(-1)^{2+1}\begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{21}} + \overbrace{3(-1)^{3+1}\begin{vmatrix} 2 & 3 \\ 3 & 2 \end{vmatrix}}^{\mathrm{cof}(A)_{31}}$$

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.

$$\det(A) = \overbrace{4(-1)^{2+1}\begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{21}} + \overbrace{3(-1)^{2+2}\begin{vmatrix} 1 & 3 \\ 3 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{22}} + \overbrace{2(-1)^{2+3}\begin{vmatrix} 1 & 2 \\ 3 & 2 \end{vmatrix}}^{\mathrm{cof}(A)_{23}}$$

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 $\det(A) = 0$.

Watch the accompanying video for a visual demonstration of the this concept.

- **Video Length**: 5 minutes 25 seconds
- **Context**: This video demonstrates how to find a 3×3 determinant using cofactor expansion: $A = \begin{bmatrix} 5 & 15 & 10 \\ -3 & -8 & -6 \\ -2 & -5 & -3 \end{bmatrix}$

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

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

$$\begin{aligned} \det(A) =& 1 \begin{vmatrix} 2 & 6 & 7 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix} + 0(-1)^{2+1} \begin{vmatrix} 2 & 3 & 77 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix} + \\ & 0(-1)^{3+1} \begin{vmatrix} 2 & 3 & 77 \\ 2 & 6 & 7 \\ 0 & 0 & -1 \end{vmatrix} + 0(-1)^{4+1} \begin{vmatrix} 2 & 3 & 77 \\ 2 & 6 & 7 \\ 0 & 3 & 33.7 \end{vmatrix} \end{aligned} \tag{3.1.1}$$

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

$$\det(A) = 1 \times \left(2 \times \begin{vmatrix} 3 & 33.7 \\ 0 & -1 \end{vmatrix} + 0(-1)^{2+1} \begin{vmatrix} 6 & 7 \\ 0 & -1 \end{vmatrix} + 0(-1)^{3+1} \begin{vmatrix} 6 & 7 \\ 3 & 33.7 \end{vmatrix}\right)$$

$$= 1 \times 2 \times \begin{vmatrix} 3 & 33.7 \\ 0 & -1 \end{vmatrix}$$

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 Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

- 3.1: Basic Techniques by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

3.1E: Exercises for Section 3.1

Exercise $3.1E.1$

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

Answer available at the end of this section.

Exercise $3.1E.2$

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

a. $\text{minor}(A)_{11}$
b. $\text{minor}(A)_{21}$
c. $\text{minor}(A)_{32}$
d. $\text{cof}(A)_{11}$
e. $\text{cof}(A)_{21}$
f. $\text{cof}(A)_{32}$

Answer available at the end of this section.

Exercise $3.1E.3$

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 available at the end of this section.

Exercise $3.1E.4$

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 available at the end of this section.

Exercise $3.1E.5$

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 available at the end of this section.

Exercise $3.1E.6$

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 available at the end of this section.

Exercise $3.1E.7$

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 available at the end of this section.

Exercise $3.1E.8$

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 $3.1E.9$

Prove that for any 2×2 or 3×3 matrix A, if one row is a multiple of another row, then the $\det(A) = 0$

Exercise $3.1E.10$

Find the value of k such that $\det(A) = 0$ for the following matrix.

$$\begin{vmatrix} k & 4 & 7 \\ -2 & 0 & 8 \\ 0 & -6 & 9 \end{vmatrix}$$

Answer available at the end of this section.

Exercise $3.1E.11$

Suppose A is a 2×2 matrix. What effect will the following have on the determinant of the resulting matrix?

a. Multiplying the first row by 2.
b. Multiplying the second column by -3.
c. Multiplying the entire matrix by 5.

Answer available at the end of this section.

Exercise $3.1E.12$

Find the value of c such that $\det(A) \neq 0$ for the following matrix.

$$\begin{vmatrix} 3 & 2 & 0 & 6 \\ -1 & 5 & 6 & c \\ 0 & -2 & -2 & 0 \\ 4 & -3 & 1 & 8 \end{vmatrix}$$

Answer available at the end of this section.

Answers

Exercise $3.1E.1$

Exercise $3.1E.2$

Exercise $3.1E.3$

a. The answer is 31.
b. The answer is 375.
c. The answer is -2.

Exercise $3.1E.4$

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix} = 6$$

Exercise $3.1E.5$

$$\begin{vmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{vmatrix} = 2$$

Exercise $3.1E.6$

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix} = 6$$

Exercise $3.1E.7$

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

Exercise $3.1E.10$

$k = -\frac{13}{4}$

Exercise $3.1E.11$

a. The determinant will be multiplied by a factor of 2.
b. The determinant will be multiplied by a factor of -3.
c. The determinant will be multiplied by a factor of 5^2.

Exercise $3.1E.12$

$c \neq -2$

This page titled 3.1E: Exercises for Section 3.1 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

3.2: Properties of Determinants

Learning Objectives

- Verify and apply the following properties:
  - $\det(AB) = \det(A)\det(B)$ (determinant of a product)
  - $\det(A^T) = \det(A)$ (determinant of a transpose)
  - $\det(A^{-1}) = \dfrac{1}{\det(A)}$ when invertible
  - $\det(kA) = k^n \det(A)$ for $n \times n$ matrices
- Demonstrate the effects that elementary row operations have on determinants.

Properties of Determinants (Core)

Before we explore how row operations affect determinants, it helps to gather the most fundamental properties first. These properties connect determinants to matrix multiplication, inverses, transposes, and scaling by constants. Together they form the "toolkit" you will repeatedly use in computations and proofs.

Theorem 3.2.5: Determinant of a Product

If A, B are $n \times n$ matrices, then $\det(AB) = \det(A)\det(B)$.

Proof (click to expand)

If A is an elementary matrix, multiplying by A on the left performs a row operation and the formula follows. For a general A, write $A = E_1 \cdots E_m C$ with C in reduced row echelon form. If $C = I$, repeated application proves the result. If C has a zero row, both $\det(A)$ and $\det(AB)$ are zero. Thus in all cases, $\det(AB) = \det(A)\det(B)$.

This property is especially important: it tells us determinants respect matrix multiplication in the same way as absolute values respect multiplication of numbers.

Theorem 3.2.6: Determinant of the Transpose

$\det(A^T) = \det(A)$.

Proof (click to expand)

For elementary matrices, $\det(E) = \det(E^T)$. Writing $A = E_1 \cdots E_m C$, we see $A^T = C^T E_m^T \cdots E_1^T$. Since $\det(E_j) = \det(E_j^T)$ and $\det(C) = \det(C^T)$, it follows $\det(A^T) = \det(A)$.

Notice that this means determinants do not depend on whether we view a matrix by rows or by columns. This symmetry will also justify why column operations behave like row operations.

Theorem 3.2.7: Determinant of the Inverse

If A is invertible, then $\det(A^{-1}) = \dfrac{1}{\det(A)}$.

Proof (click to expand)

Since $AA^{-1} = I$, $\det(A)\det(A^{-1}) = \det(I) = 1$. Hence $\det(A^{-1}) = 1/\det(A)$.

This connects determinants to invertibility: a matrix is invertible exactly when its determinant is nonzero. Thus, determinants provide a quick "test for invertibility."

Theorem 3.2.3: Scalar Multiplication

If $B = kA$, then $\det(B) = k^n \det(A)$.

Proof (click to expand)

Multiplying each row of A by k multiplies the determinant by k. Doing this for all n rows multiplies by k^n, so $\det(kA) = k^n \det(A)$.

Together, these four properties give us a powerful set of shortcuts when evaluating determinants or analyzing matrices. Next we turn to row operations, where we see how basic manipulations affect determinant values.

Row Operations and Determinants

In Chapter 1, row operations were used to simplify matrices and solve systems. Now we revisit them with a new perspective: every row operation has a predictable effect on the determinant. Understanding this allows us to compute determinants of large matrices more efficiently.

Theorem (Row Operations)

1. Swapping two rows multiplies the determinant by -1.
2. Multiplying a row by k multiplies the determinant by k.
3. Adding a multiple of one row to another leaves the determinant unchanged.

Proof (click to expand)

Each case is checked directly for 2×2 matrices, then extended by induction. For example, if two rows are identical, interchanging them shows $\det A = -\det A$, so $\det A = 0$. Multiplying a row scales the determinant by that factor. Adding a multiple of a row to another preserves cofactors, so the determinant does not change.

These rules mean that when simplifying a matrix using row operations, you can keep track of how the determinant changes without having to recompute it from scratch.

Example: Switching Rows

$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $\det(A) = -2$. Swapping rows $\rightarrow$ $\det(B) = 2$.

Example: Adding a Multiple of a Row

$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 \\ 5 & 8 \end{bmatrix}$. Row2(B) = Row2(A)+2·Row1(A). $\det(B) = \det(A) = -2$.

With these principles in place, we can now attempt larger matrices. The goal is to apply row operations to simplify the matrix, carefully tracking determinant changes, until the determinant can be computed by expansion or by recognizing a triangular form.

Example 3.2.10: Large Determinant

For $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 1 & 2 & 3 \\ 4 & 5 & 4 & 3 \\ 2 & 2 & -4 & 5 \end{bmatrix}$, row ops lead to an upper-triangular matrix with product of diagonal entries 1485. Adjusting for scaling, $\det(A) = 495$.

This example shows how quickly row operations reduce the amount of work compared to Laplace expansion.

Example 3.2.11

For $A = \begin{bmatrix} 1 & 2 & 3 & 2 \\ 1 & -3 & 2 & 1 \\ 2 & 1 & 2 & 5 \\ 3 & -4 & 1 & 2 \end{bmatrix}$, careful row and column operations simplify it to a form where expansion gives $\det(A) = -82$.

Row operations provide a powerful computational strategy, and combining them with the determinant properties from earlier makes it possible to handle both theory and computation with confidence.

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 Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

- 3.2: Properties of Determinants by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

3.2E: Exercises for Section 3.2

Exercise 3.2E.1

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 available at the end of this section.

Exercise 3.2E.2

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 available at the end of this section.

Exercise 3.2E.3

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 available at the end of this section.

Exercise 3.2E.4

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 available at the end of this section.

Exercise 3.2E.5

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 available at the end of this section.

Exercise $3.2E.6$

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 available at the end of this section.

Exercise $3.2E.7$

Show $\det(aA) = a^n \det(A)$ for an $n \times n$ matrix A and scalar a.

Answer available at the end of this section.

Exercise $3.2E.8$

Construct 2×2 matrices A and B to show that the $\det A \det B = \det(AB)$.

Answer available at the end of this section.

Exercise $3.2E.9$

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 available at the end of this section.

Exercise $3.2E.10$

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 available at the end of this section.

Exercise $3.2E.11$

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 available at the end of this section.

Exercise $3.2E.12$

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 available at the end of this section.

Exercise $3.2E.13$

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 available at the end of this section.

Exercise $3.2E.14$

Find the determinant using row operations to first simplify.

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

Answer available at the end of this section.

Exercise $3.2E.15$

Find the determinant using row operations to first simplify.

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

Answer available at the end of this section.

Exercise $3.2E.16$

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 available at the end of this section.

Exercise $3.2E.17$

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 available at the end of this section.

Answers

Exercise $3.2E.1$

It does not change the determinant. This was just taking the transpose.

Exercise $3.2E.2$

In this case two rows were switched and so the resulting determinant is -1 times the first

Exercise $3.2E.3$

The determinant is unchanged. It was just the first row added to the second.

Exercise $3.2E.4$

The second row was multiplied by 2 so the determinant of the result is 2 times the original determinant.

Exercise $3.2E.5$

In this case the two columns were switched so the determinant of the second is -1 times the determinant of the first.

Exercise $3.2E.6$

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 $3.2E.7$

$\det(aA) = \det(aIA) = \det(aI)\det(A) = a^n \det(A)$. The matrix which has a down the main diagonal has determinant equal to a^n.

Exercise $3.2E.8$

$$\det\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} -1 & 2 \\ -5 & 6 \end{bmatrix}\right) = -8$$

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

Exercise $3.2E.9$

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 $3.2E.10$

It must be 0 because $0 = \det(0) = \det(A^k) = (\det(A))^k$.

Exercise $3.2E.11$

You would need $\det(AA^T) = \det(A)\det(A^T) = \det(A)^2 = 1$ and so $\det(A) = 1$, or -1.

Exercise $3.2E.12$

$\det(A) = \det(S^{-1}BS) = \det(S^{-1})\det(B)\det(S) = \det(B)\det(S^{-1}S) = \det(B)$.

Exercise $3.2E.13$

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 $3.2E.14$

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

Exercise $3.2E.15$

$$\begin{vmatrix} 2 & 1 & 3 \\ 2 & 4 & 2 \\ 1 & 4 & -5 \end{vmatrix} = -32$$

Exercise $3.2E.16$

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.

$$\begin{vmatrix} 1 & 2 & 1 & 2 \\ 3 & 1 & -2 & 3 \\ -1 & 0 & 3 & 1 \\ 2 & 3 & 2 & -2 \end{vmatrix} = 63$$

Exercise $3.2E.17$

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

$$\begin{vmatrix} 1 & 4 & 1 & 2 \\ 3 & 2 & -2 & 3 \\ -1 & 0 & 3 & 3 \\ 2 & 1 & 2 & -2 \end{vmatrix} = 211$$

This page titled 3.2E: Exercises for Section 3.2 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

3.3: Application of the Determinant to Inverses; Cramer's Rule

Learning Objectives

- Use determinants to determine whether a matrix has an inverse, and evaluate the inverse using cofactors.
- Apply Cramer's Rule to solve a 2×2 or a 3×3 linear system.

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 2.2.1. 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 3.3.1: 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 $\mathrm{adj}(A)$.

In the specific case where A is a 2×2 matrix given by

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

then $\mathrm{adj}(A)$ is given by

$$\mathrm{adj}(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

In general, $\mathrm{adj}(A)$ 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 3.3.1: The Inverse and the Determinant

Let A be an $n \times n$ matrix. Then

$$A\,\mathrm{adj}(A) = \mathrm{adj}(A)\,A = \det(A) I$$

Moreover A is invertible if and only if $\det(A) \neq 0$. In this case we have:

$$A^{-1} = \frac{1}{\det(A)} \mathrm{adj}(A)$$

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}.

Before proving the theorem, let's consider the following example.

Example 3.3.1: 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 3.3.1.

Solution

According to Theorem 3.3.1,

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A)$$

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 $\operatorname{adj}(A)$. To do so, first we will find the cofactor matrix of A. This is given by

$$\operatorname{cof}(A) = \begin{bmatrix} -2 & -2 & 6 \\ 4 & -2 & 0 \\ 2 & 8 & -6 \end{bmatrix}$$

Here, the ij^{th} entry is the ij^{th} cofactor of the original matrix A which you can verify. Therefore, from Theorem 3.3.1, 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 3.3.2: 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 3.3.1.

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

$$A^{-1} = \frac{1}{(1/6)} \operatorname{adj}(A) = 6 \operatorname{adj}(A)$$

We continue to calculate as follows. Here we show the 2×2 determinants needed to find the cofactors.

$$A^{-1} = 6 \begin{bmatrix} \begin{vmatrix} \frac{1}{3} & -\frac{1}{2} \\ \frac{2}{3} & -\frac{1}{2} \end{vmatrix} & - \begin{vmatrix} -\frac{1}{6} & -\frac{1}{2} \\ -\frac{5}{6} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} -\frac{1}{6} & \frac{1}{3} \\ -\frac{5}{6} & \frac{2}{3} \end{vmatrix} \\ - \begin{vmatrix} 0 & \frac{1}{2} \\ \frac{2}{3} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{5}{6} & -\frac{1}{2} \end{vmatrix} & - \begin{vmatrix} \frac{1}{2} & 0 \\ -\frac{5}{6} & \frac{2}{3} \end{vmatrix} \\ \begin{vmatrix} 0 & \frac{1}{2} \\ \frac{1}{3} & -\frac{1}{2} \end{vmatrix} & - \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{6} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} \frac{1}{2} & 0 \\ -\frac{1}{6} & \frac{1}{3} \end{vmatrix} \end{bmatrix}^T$$

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 restate and prove Theorem 3.3.1.

Theorem 3.3.1

Let A be an $n \times n$ matrix. Then

$$A \operatorname{adj}(A) = \operatorname{adj}(A) A = \det(A) I$$

Moreover A is invertible if and only if $\det(A) \neq 0$. In this case we have:

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A)$$

Proof available at the end of this section.

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 3.3.3: 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 3.3.1 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,

$$\begin{aligned} AX &= B \\ A^{-1}(AX) &= A^{-1}B \\ (A^{-1}A)X &= A^{-1}B \\ IX &= A^{-1}B \\ X &= A^{-1}B \end{aligned} \tag{3.3.1}$$

Hence, the solutions X to the system are given by $X = A^{-1}B$. Since we assume that A^{-1} exists, we can use the formula for A^{-1} given above. Substituting this formula into the equation for X, we have

$$X = A^{-1}B = \frac{1}{\det(A)}\operatorname{adj}(A)B$$

Let x_i be the ith entry of X and b_j be the jth entry of B. Then this equation becomes

$$x_i = \sum_{j=1}^{n} [a_{ij}]^{-1} b_j = \sum_{j=1}^{n} \frac{1}{\det(A)}\operatorname{adj}(A)_{ij} b_j$$

where $\operatorname{adj}(A)_{ij}$ is the ij^{th} entry of $\operatorname{adj}(A)$.

By the formula for the expansion of a determinant along a column,

$$x_i = \frac{1}{\det(A)} \det \begin{bmatrix} * & \cdots & b_1 & \cdots & * \\ \vdots & & \vdots & & \vdots \\ * & \cdots & b_n & \cdots & * \end{bmatrix}$$

where here the 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 3.3.1: 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 3.3.4: 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 3.3.1 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, the column corresponding to x, 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, the column corresponding to y, 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, the column corresponding to z, 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.

Example 3.3.5: Use Cramer's Rule for Non-Constant Matrix

Solve for z if

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix}$$

Solution

We are asked to find the value of z in the solution. We will solve using Cramer's rule. Thus

$$z = \frac{\begin{vmatrix} 1 & 0 & 1 \\ 0 & e^t \cos t & t \\ 0 & -e^t \sin t & t^2 \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{vmatrix}} = t\left(\left(\cos t\right) t + \sin t\right) e^{-t}$$

Answers

Theorem 3.3.1

Recall that the (i,j)-entry of $\mathrm{adj}(A)$ is equal to $\mathrm{cof}(A)_{ji}$. Thus the (i,j)-entry of $B = A \cdot \mathrm{adj}(A)$ is :

$$B_{ij} = \sum_{k=1}^{n} a_{ik}\mathrm{adj}(A)_{kj} = \sum_{k=1}^{n} a_{ik}\mathrm{cof}(A)_{jk}$$

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:

$$A \,\mathrm{adj}\,(A) = \det(A) I$$

Similarly we can verify that:

$$\mathrm{adj}\,(A)\, A = \det(A) I$$

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:

$$A^{-1} = \frac{1}{\det(A)}\mathrm{adj}\,(A)$$

This completes the proof.

This page titled 3.3: Application of the Determinant to Inverses; Cramer's Rule is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson 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.

3.3E: Exercises for Section 3.3

Exercise 3.3E.1

Determine whether the following matrices are invertible by computing their determinants.

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

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

Answer available at the end of this section.

Exercise 3.3E.2

Find the adjugate of the matrix:

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

Answer available at the end of this section.

Exercise 3.3E.3

Find the inverse of the matrix

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

using the determinant and the adjugate matrix.

Answer available at the end of this section.

Exercise 3.3E.4

Find the inverse of the matrix

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

using the determinant and the adjugate matrix

Exercise 3.3E.5

Find the cofactor matrix for:

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

Exercise 3.3E.6

Find the inverse of

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

by computing the cofactor matrix and adjugate.

Exercise 3.3E.7

Solve the system

$$\begin{cases} 2x + 3y = 5 \\ 4x + y = 6 \end{cases}$$

using Cramer's Rule.

Answer available at the end of this section.

Exercise 3.3E.8

Solve the system

$$\begin{cases} x + 2y + 3z = 6 \\ 2x - y + z = 3 \\ 3x + y - 2z = 4 \end{cases}$$

using Cramer's Rule

Exercise 3.3E.9

Find a 3×3 system where Cramer's Rule does not apply, and explain why.

Hint available at the end of this section.

Exercise 3.3E.10

Show that for any invertible 2×2 matrix A,

$$A^{-1} = \frac{1}{\det(A)} \text{adj}(A).$$

Exercise 3.3E.11

If A is an upper triangular 3×3 matrix, explain how Cramer's Rule simplifies solving $Ax = b$.

Exercise 3.3E.12

In an electrical circuit, Kirchhoff's Laws lead to the system:

$$\begin{cases} 3I_1 + I_2 - I_3 = 5 \\ 2I_1 + 4I_2 + I_3 = 6 \\ -I_1 + 2I_2 + 5I_3 = 2 \end{cases}$$

Solve for the currents I_1, I_2, I_3 using Cramer's Rule.

Answer available at the end of this section.

Exercise $3.3E.13$

If the determinant of a matrix is very small but nonzero, what does this tell us about the matrix?

Answer available at the end of this section.

Exercise $3.3E.14$

Solve the system of equations using Cramer's Rule:

$$\begin{cases} x + y + z + w = 4 \\ 2x - y + 3z + w = 5 \\ x + 2y - z - w = 2 \\ 3x + y + 2z + 4w = 8 \end{cases}$$

Answer available at the end of this section.

Exercise $3.3E.15$

Consider the system:

$$\begin{cases} ax + y = 3 \\ 2x + (a+1)y = 5 \end{cases}$$

a. For what values of a does the system have no unique solution?
b. For what values of a does the system have a unique solution, and what is it in terms of a?

Answers

Exercise $3.3E.1$

a. $\det(A) = 5$. Invertible.
b. $\det(B) = 0$. Not invertible.

Exercise $3.3E.2$

The cofactor matrix is:

$$\text{cof}(A) = \begin{bmatrix} 1 & -2 \\ -3 & 4 \end{bmatrix}$$

The adjugate (adjoint) matrix is the transpose of the cofactor matrix:

$$\text{adj}(A) = \begin{bmatrix} 1 & -3 \\ -2 & 4 \end{bmatrix}$$

Exercise $3.3E.3$

First, compute the determinant:

$$\det(A) = (1)(4) - (2)(3) = 4 - 6 = -2.$$

Next, compute the cofactor matrix:

$$\text{cof}(A) = \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}$$

Now, compute the adjugate (transpose of the cofactor matrix):

$$\text{adj}(A) = \begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}$$

Using the formula for the inverse:

$$A^{-1} = \frac{1}{\det(A)}\text{adj}(A),$$

we substitute the values:

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

Thus, the inverse is:

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

Exercise $3.3E.7$

$x = \frac{\det(A_x)}{det(A)}, y = \frac{\det(A_y)}{det(A)}$, yielding $x = 1, y = 1$

Exercise $3.3E.9$

What must be true about determinant of the coefficient matrix for Cramer's Rule to work?

Exercise $3.3E.12$

Using determinants, $I_1 = 1, I_2 = 2, I_3 = -1$.

Exercise $3.3E.13$

There are at least two reasons:

1. While the matrix technically has an inverse, numerical computations involving the inverse may be unstable due to round-off errors.
2. Even small changes in input (such as rounding errors) can cause large errors in the output when solving systems of equations.

Exercise $3.3E.14$

$x = 2, y = -1, z = 3, w = 4$

This page titled 3.3E: Exercises for Section 3.3 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

3.4: Determinants and Geometry

Learning Objectives

- Understand the relationship between the determinant of a matrix and the volume of a parallelepiped.
- Learn to use determinants to compute areas of parallelograms and triangles.

In this section we give a geometric interpretation of determinants, in terms of *areas* and *volumes*.

Recall the three properties of the determinants related to the elementary row operations. For an $n \times n$ matrix A:

1. If B is a matrix that results from switching two rows of A, then $\det(B) = -\det(A)$.
2. If B is a matrix that results from multiplying some row of A by a scalar k, then $\det(B) = k\det(A)$.
3. If B is a matrix that results from adding a multiple of a row of A to another row of A, then $\det(B) = \det(A)$.

Additionally, note that the determinant of an $n \times n$ identify matrix I_n is 1.

Parallelograms and Parallelepipeds

The determinant computes the volume of the following kind of geometric object.

Definition 3.4.1: Parallelepiped

The ***parallelepiped*** determined by n vectors $\vec{v}_1, \vec{v}_, \dots, \vec{v}_n$ in $\mathbb{R}^n$ is the subset

$$P = \left\{ a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n \;\middle|\; 0 \le a_1, a_2, \dots, a_n \le 1 \right\}.$$

In other words, a parallelepiped is the set of all linear combinations of n vectors with coefficients in $[0, 1]$. We can draw parallelepipeds using the parallelogram law for vector addition.

Example 3.4.1: The Unit Cube

The parallelepiped determined by the standard coordinate vectors $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$ is the unit n-dimensional cube.

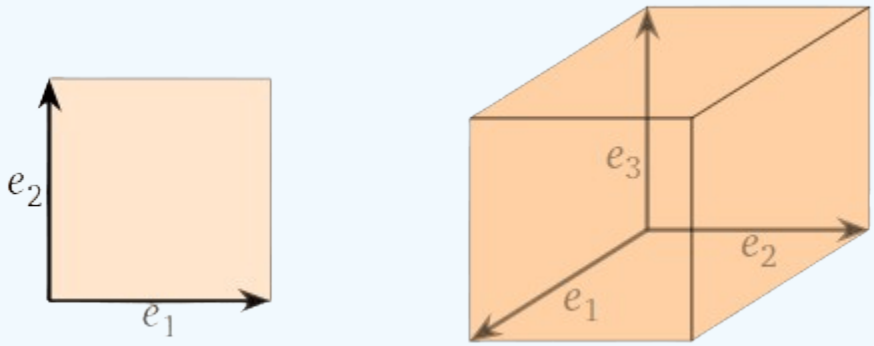


Figure 3.4.1: A 2D square with axes $\vec{e}_1$ and $\vec{e}_2$ on the left, and a 3D cube with axes $\vec{e}_1$, $\vec{e}_2$, and $\vec{e}_3$ on the right, representing coordinate systems. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

Example 3.4.2: Parallelograms

When $n = 2$, a parallelepiped is just a parallelogram in $\mathbb{R}^2$. Note that the edges come in parallel pairs.

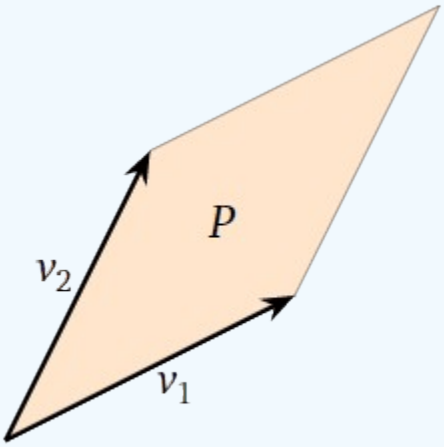


Figure 3.4.2: A parallelogram P formed by vectors $\vec{v}_1$ and $\vec{v}_2$, represented as black arrows from a common point. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

Example 3.4.3

When $n = 3$, a parallelepiped is a kind of a skewed cube. Note that the faces come in parallel pairs.

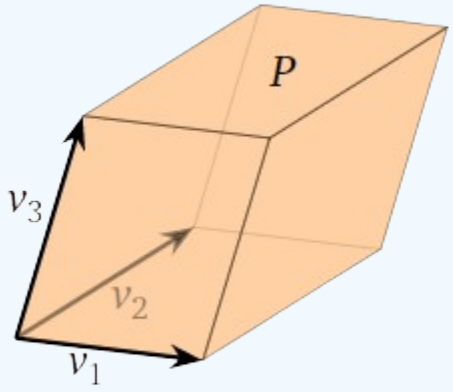


Figure 3.4.3: A parallelepiped P formed by vectors $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_3$, represented as black arrows from a common point. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

When does a parallelepiped have zero volume? This can happen only if the parallelepiped is flat, i.e., it is squashed into a lower dimension.

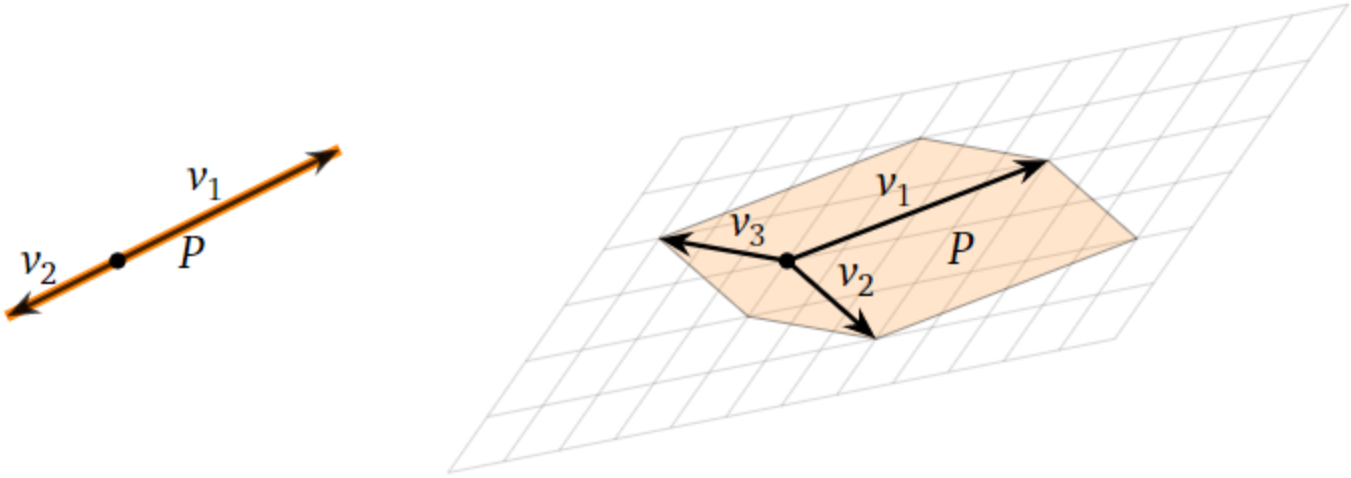


Figure 3.4.4: An illustration comparing geometric objects in one and two dimensions. On the left, a one-dimensional space is represented by a line segment with two vectors, $\vec{v}_1$ and $\vec{v}_2$, extending in opposite directions from a point P. On the right, a two-dimensional parallelogram is shown lying in a tilted plane on a grid. The parallelogram is defined by three vectors, $\vec{v}_1, \vec{v}_2$, and $\vec{v}_3$, originating from a common point P and forming a shaded area. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

This means exactly that the set $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ has some dependence, implying that the matrix with rows $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ has determinant zero. To summarize:

Key Observation 3.4.1

The parallelepiped defined by $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ has zero volume if and only if the matrix with rows $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ has zero determinant.

Determinants and Volumes

The key observation above is only the beginning of the story: the volume of a parallelepiped is *always* a determinant.

Theorem 3.4.1: Determinants and Volumes

Let $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ be vectors in $\mathbb{R}^n$, let P be the parallelepiped determined by these vectors, and let A be the matrix with rows $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$. Then the absolute value of the determinant of A is the volume of P:

$$|\det(A)| = \text{vol}(P).$$

Proof available at the end of this section.

Note

Since $\det(A) = \det(A^T)$, the absolute value of $\det(A)$ is also equal to the volume of the parallelepiped determined by the *columns* of A as well.

Example 3.4.4: Length

A 1×1 matrix A is just a number $\begin{pmatrix} a \end{pmatrix}$. In this case, the parallelepiped P determined by its one row is just the interval $[0, a]$ (or $[a, 0]$ if $a < 0$). The "volume" of a region in $\mathbb{R}^1 = \mathbb{R}$ is just its length, so it is clear in this case that $\text{vol}(P) = |a|$.

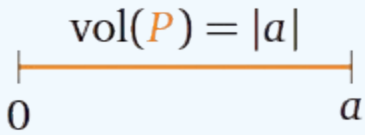


Figure 3.4.8: A diagram illustrating the volume (length) of a one-dimensional interval. A horizontal orange line segment extends from 0 to a, with vertical bars at both endpoints. Above the segment, the equation $\text{vol}(P) = |a|$ is displayed, indicating that the length of the interval is given by the absolute value of a. The endpoints are labeled 0 and a, representing the interval's boundaries. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

Example 3.4.5: Area

When A is a 2×2 matrix, its rows determine a parallelogram in $\mathbb{R}^2$. The "volume" of a region in $\mathbb{R}^2$ is its area, so we obtain a formula for the area of a parallelogram: it is the determinant of the matrix whose rows are the vectors forming two adjacent sides of the parallelogram.

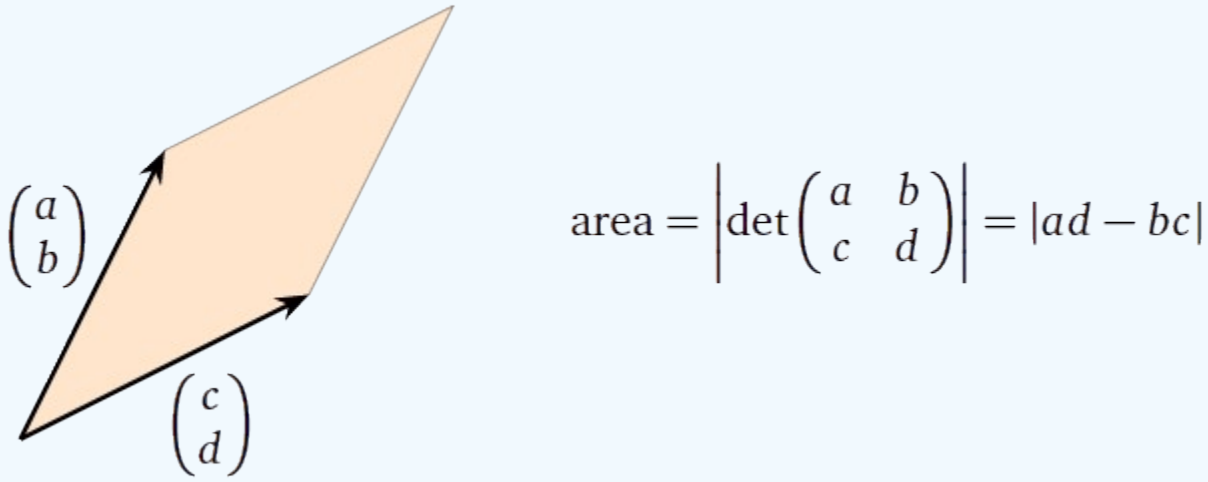


Figure 3.4.9: A diagram illustrating the area of a parallelogram using the determinant. On the left, a parallelogram is shown in a two-dimensional plane, defined by two black vectors originating from a common point. These vectors are labeled as column vectors $\begin{pmatrix} a \\ b \end{pmatrix}$ and $\begin{pmatrix} c \\ d \end{pmatrix}$. The parallelogram is shaded in light orange. On the right, the formula for the area of the parallelogram is displayed: $\text{area} = \left|\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}\right| = |ad - bc|$ which shows that the absolute value of the determinant of the matrix formed by the two vectors gives the area of the parallelogram. This diagram visually demonstrates how the determinant provides a measure of area in two dimensions. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

It is perhaps surprising that it is possible to compute the area of a parallelogram without trigonometry. It is a fun geometry problem to prove this formula by hand. [Hint: first think about the case when the first row of A lies on the x-axis.]

Example 3.4.6

Find the area of the parallelogram with sides $(1, 3)$ and $(2, -3)$.

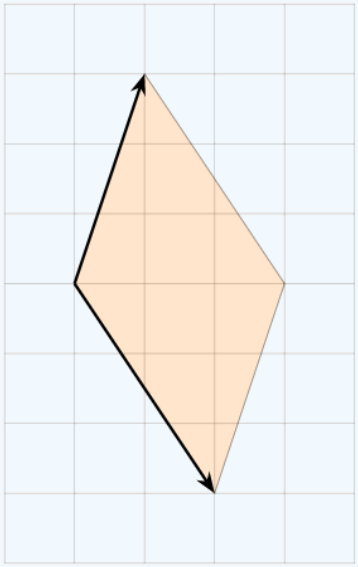


Figure 3.4.10: A parallelogram plotted on a coordinate grid, defined by two black arrows representing vectors originating from a common point. The parallelogram is shaded in light orange, illustrating the region whose area is determined by the determinant of the matrix formed by the vectors. The grid provides a reference for the positioning of the parallelogram in two-dimensional space. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

Solution

The area is

$$\left|\det\begin{pmatrix}1 & 3\\ 2 & -3\end{pmatrix}\right| = |-3-6| = 9.$$

Example 3.4.7

Find the area of the parallelogram in the picture.

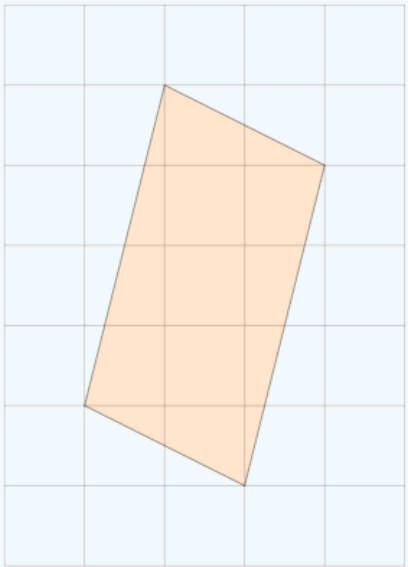


Figure 3.4.11: A parallelogram plotted on a coordinate grid. The points are not labeled. Starting with the top-left point as the first point and moving clockwise, the second point is 2 units right and 1 unit down from the first point. The third point is 1 unit left and 4 units down from the second point. The fourth point is 2 units left and 1 unit up from the third point. The first point is 1 unit right and 4 units up from the fourth point. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

Solution

We choose two adjacent sides to be the rows of a matrix. We choose the top two:

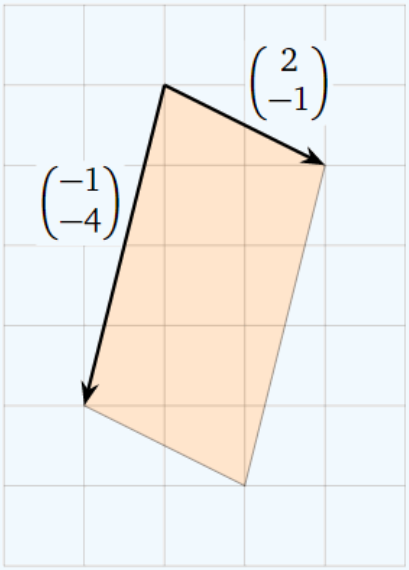


Figure 3.4.12: A parallelogram plotted on a coordinate grid, defined by two vectors originating from a common point (the top-left point of the parallelogram). The vectors are labeled with their components: $\begin{pmatrix}-1\\-4\end{pmatrix}$ and $\begin{pmatrix}2\\-1\end{pmatrix}$. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

Note that we do not need to know where the origin is in the picture: vectors are determined by their length and direction, not where they start. The area is

$$\left|\det\begin{pmatrix}-1 & -4\\ 2 & -1\end{pmatrix}\right| = |1+8| = 9.$$

Example 3.4.8: Area of a triangle

Find the area of the triangle with vertices $(-1,-2),\ (2,-1),\ (1,3)$.

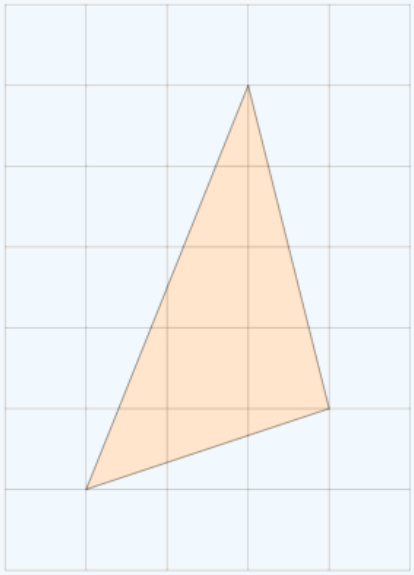


Figure 3.4.13: A triangle plotted on a coordinate grid. The triangle is formed by three points, creating a geometric region in two-dimensional space. The points are not labeled. Starting with the bottom-left point as the first point and moving clockwise, the second point is 2 units right and 5 units up. The third point is 1 unit right and 4 units down from the second point. The first point is 3 units left and 1 unit down from the third point. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

Solution

Doubling a triangle makes a parallelogram. We choose two of its sides to be the rows of a matrix.

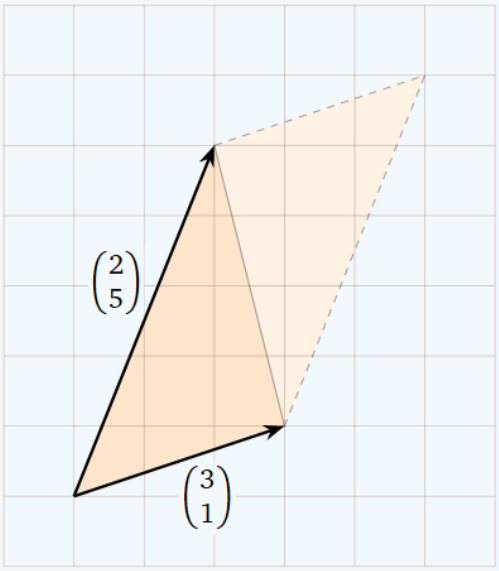


Figure 3.4.14: A triangle given in the example is extended to a parallelogram defined by two vectors originating from a common point. The vectors are labeled with their components: $\begin{pmatrix}2\\5\end{pmatrix}$ and $\begin{pmatrix}3\\1\end{pmatrix}$. The original triangle is shaded darker than the triangle added to form the parallelogram. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

The area of the parallelogram is

$$\left|\det\begin{pmatrix}2&5\\3&1\end{pmatrix}\right| = |2-15| = 13,$$

so the area of the triangle is $13/2$.

You might be wondering: if the absolute value of the determinant is a volume, what is the geometric meaning of the determinant without the absolute value? The next remark explains that we can think of the determinant as a *signed* volume. If you have taken an integral calculus course, you probably computed negative areas under curves; the idea here is similar.

Remark: Signed volumes

Theorem 3.4.1 on determinants and volumes tells us that the *absolute value* of the determinant is the volume of a paralellepiped. This raises the question of whether the sign of the determinant has any geometric meaning.

A 1×1 matrix A is just a number $\begin{pmatrix} a \end{pmatrix}$. In this case, the parallelepiped P determined by its one row is just the interval $[0, a]$ if $a \geq 0$, and it is $[a, 0]$ if $a < 0$. In this case, the sign of the determinant determines whether the interval is to the left or the right of the origin.

For a 2×2 matrix with rows v_1, v_2, the sign of the determinant determines whether v_2 is counterclockwise or clockwise from v_1. That is, if the counterclockwise angle from v_1 to v_2 is less than 180°, then the determinant is positive; otherwise it is negative (or zero).

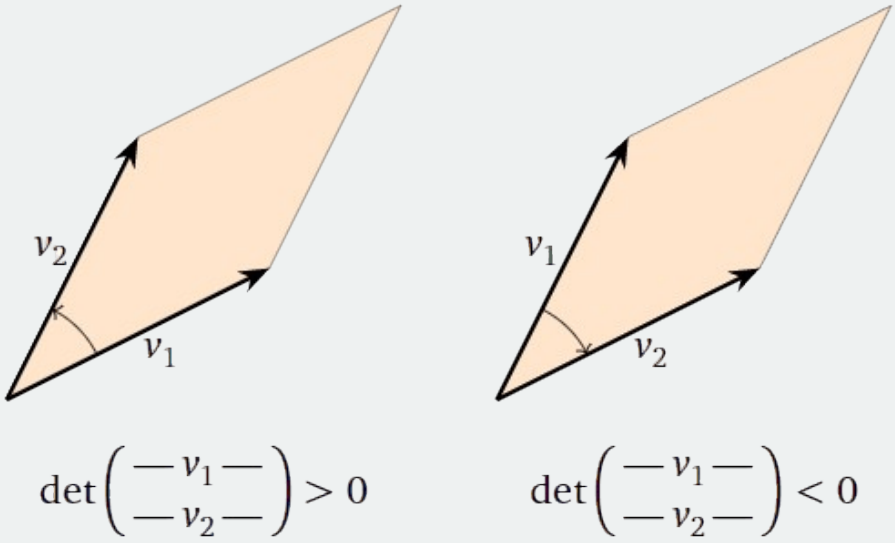


Figure 3.4.15: A comparison of two parallelograms, each defined by two vectors, $\vec{v}_1$ and $\vec{v}_2$, originating from a common point. On the left, the parallelogram is oriented such that the determinant of the matrix formed by $\vec{v}_1$ and $\vec{v}_2$ is positive, indicated by the expression: $\det(\vec{v}_1 \quad \vec{v}_2) > 0$. A signed angle between the vectors is shown and is positive, emphasizing their orientation. On the right, the parallelogram is oriented in the opposite way, resulting in a negative determinant: $\det(\vec{v}_1 \quad \vec{v}_2) < 0$. The reversed orientation demonstrates the effect of swapping the vectors on the determinant's sign. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

For example, if $v_1 = \begin{pmatrix} a \\ b \end{pmatrix}$, then the counterclockwise rotation of v_1 by 90° is $v_2 = \begin{pmatrix} -b \\ a \end{pmatrix}$ by Example 3.3.8 in Section 3.3, and

$$\det\begin{pmatrix} a & b \\ -b & a \end{pmatrix} = a^2 + b^2 > 0.$$

On the other hand, the *clockwise* rotation of v_1 by 90° is $\begin{pmatrix} b \\ -a, \end{pmatrix}$ and

$$\det\begin{pmatrix} a & b \\ b & -a \end{pmatrix} = -a^2 - b^2 < 0.$$

For a 3×3 matrix with rows v_1, v_2, v_3, the *right-hand rule* determines the sign of the determinant. If you point the index finger of your right hand in the direction of v_1 and your middle finger in the direction of v_2, then the determinant is positive if your thumb points roughly in the direction of v_3, and it is negative otherwise.

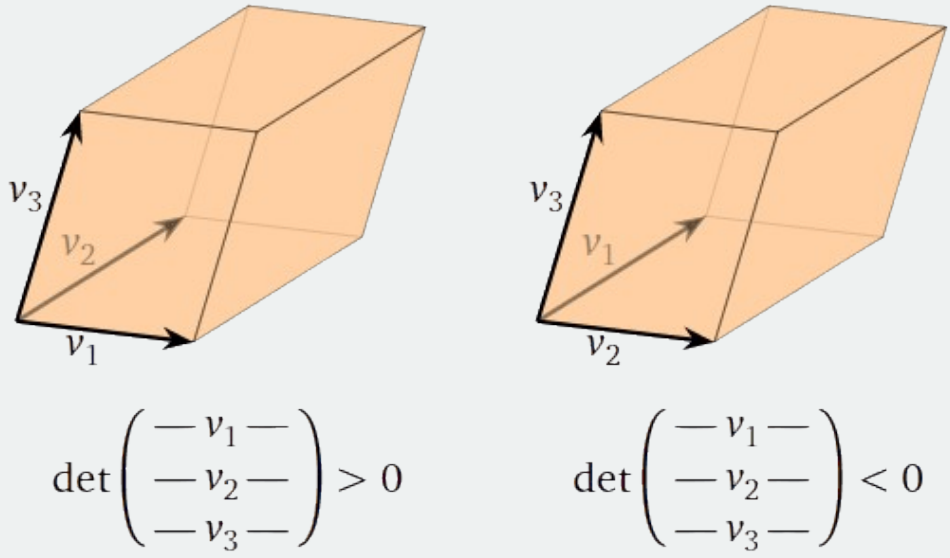


Figure 3.4.16: A comparison of two three-dimensional parallelepipeds, each defined by three vectors, $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_3$, originating from a common point. On the left, the parallelepiped is oriented such that the determinant of the matrix formed by $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_3$ is positive, indicated by the expression: $\det(\vec{v}_1 \quad \vec{v}_2 \quad \vec{v}_3) > 0$. The vectors are labeled and shown extending from a common vertex. On the right, the parallelepiped is oriented differently, resulting in a negative determinant: $\det(\vec{v}_1 \quad \vec{v}_2 \quad \vec{v}_3) < 0$. The reversed orientation demonstrates the effect of changing the order of the vectors on the determinant's sign. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

In higher dimensions, the notion of signed volume is still important, but it is usually *defined* in terms of the sign of a determinant.

Answers

Theorem 3.4.1: Determinants and Volumes

Case 1: There is a dependence in the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$. In this case, we established that the volume of the parallelopiped is 0 in Key Observation 3.4.1.

Case 2: There is no dependence in the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$. Hence the matrix A must row reduce to I_n. Suppose the required row operations correspond to left multiplying the row of A by elementary matrices $E_1, E_2, \ldots, E_k$. That is

$$E_m \ldots E_2 E_1 A = I_n$$

Since all elementary matrices are invertible and are elementary matrices themselves, we can express the above relationship as

$$A = E_1^{-1} E_2^{-1} \ldots E_m^{-1} I_n$$

Let's begin with the $n \times n$ identity matrix I_n. This matrix consists of row vectors $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$. The volume of the parallelepiped formed by these vectors is 1.

Now let us apply the opposite of the elementary row operations in reverse used to row reduce A to I_n. This would give us the row operations needed to transform I_n to A using elementary row operations. What effect do these row operations have on the absolute value of the determinant?

We begin with I_n, and note that $|\det(I_n)| = 1$.

Suppose the next row operation consists of switching two rows. Switching two rows of a matrix just reorders the vectors $v_1, v_2, \ldots, v_n$, hence has no effect on the parallelepiped determined by those vectors. Thus, the volume is unchanged by row swaps. The determinant after swapping rows will be negative, but the *absolute value* of the determinant is unchanged.

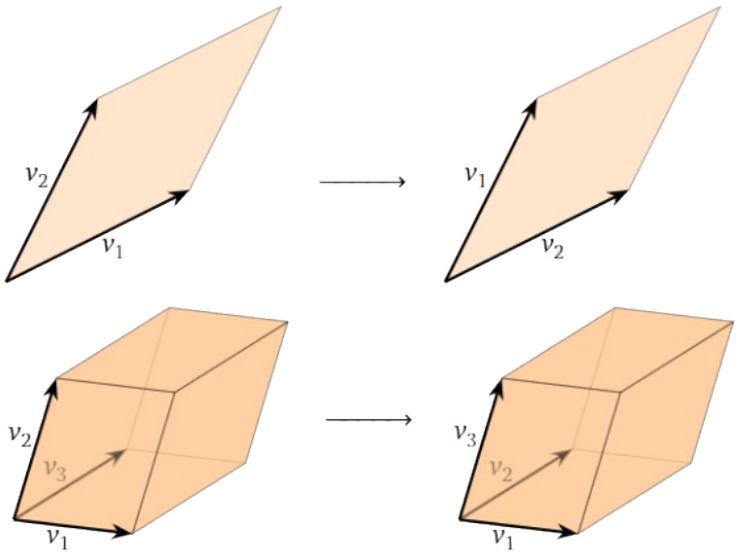


Figure 3.4.5: A diagram illustrating the effect of row swapping on area and volume in linear algebra. In the top row, two parallelograms are shown. On the left, the parallelogram is defined by vectors $\vec{v}_1$ and $\vec{v}_2$, with $\vec{v}_1$ positioned as the first vector and $\vec{v}_2$ as the second. On the right, the parallelogram remains the same in shape and size, but the roles of $\vec{v}_1$ and $\vec{v}_2$ are swapped. This transformation demonstrates that swapping two rows negates the determinant. In the bottom row, two parallelepipeds are shown. On the left, the parallelepiped is defined by vectors $\vec{v}_1, \vec{v}_2$, and $\vec{v}_3$, with the volume determined by their orientation. On the right, $\vec{v}_2$ and $\vec{v}_3$ have swapped places, maintaining the shape but flipping the determinant's sign. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

Now , suppose the next row operation consists of multiplying a row by a constant k. For simplicity, we consider a row scaling of the form $kR_n -> R_n$. This scales the length of $\vec{v}_n$ by a factor of $|k|$, which also scales the perpendicular distance of $\vec{v}_n$ from the base by a factor of $|k|$. Thus, the volume of the resulting parallelepiped is scaled by $|k|$, as is the determinant of the matrix made with the tranformed vectors as rows. >

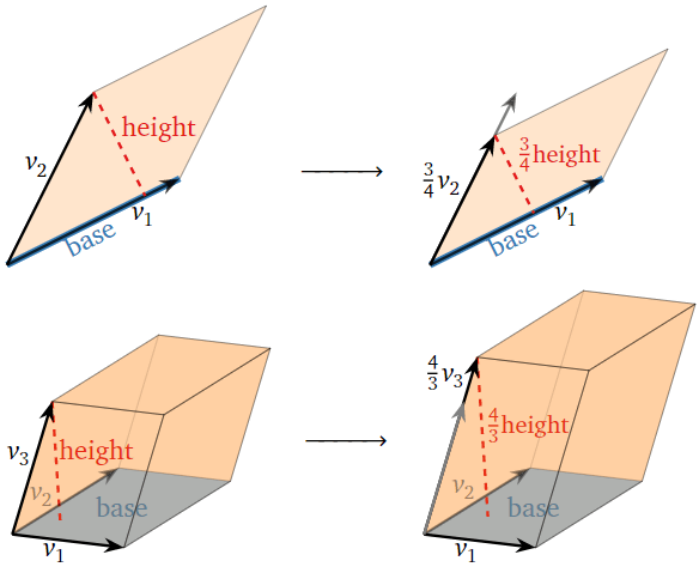


Figure 3.4.6: A diagram illustrating the effect of row scaling on area and volume in linear algebra. In the top row, two parallelograms are shown. On the left, a parallelogram is spanned by vectors $\vec{v}_1$ and $\vec{v}_2$, where $\vec{v}_1$ is labeled as the "base" in blue, and a red dashed line represents the "height" perpendicular to it. On the right, the vector v_2 is scaled by $\frac{3}{4}$, reducing the height proportionally to $\frac{3}{4}$ of its original value. This transformation illustrates that scaling a vector changes the determinant by the same factor. In the bottom row, two parallelepipeds are shown. On the left, the parallelepiped is defined by vectors $\vec{v}_1, \vec{v}_2$, and $\vec{v}_3$, where $\vec{v}_1$ and $\vec{v}_2$ form the "base" (shaded in gray), and a red dashed line represents the "height" along $\vec{v}_3$. On the right, the vector $\vec{v}_3$ is scaled by $\frac{4}{3}$, increasing the height proportionally to $\frac{4}{3}$ of its original value. This transformation illustrates that scaling a vector changes the volume by the same factor. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

Now, suppose the next row operation consists of adding a multiple of one row to another. For simplicity, we consider a row replacement of the form $R_n + cR_i -> R_n$. The volume of a parallelepiped is the volume of its base, times its height: Here the “base” is the parallelepiped determined by $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{n-1}$, and the “height” is the perpendicular distance of $\vec{v}_n$ from the base. Translating $\vec{v}_n$ by a multiple of $\vec{v}_i$ moves $\vec{v}_n$ in a direction parallel to the base. This changes neither the base nor the height! Thus, the volume of the parallelepiped is unchanged by row replacements. And as we know from before, this row operation keeps the determinant of the matrix made of the updated rows unchanged.

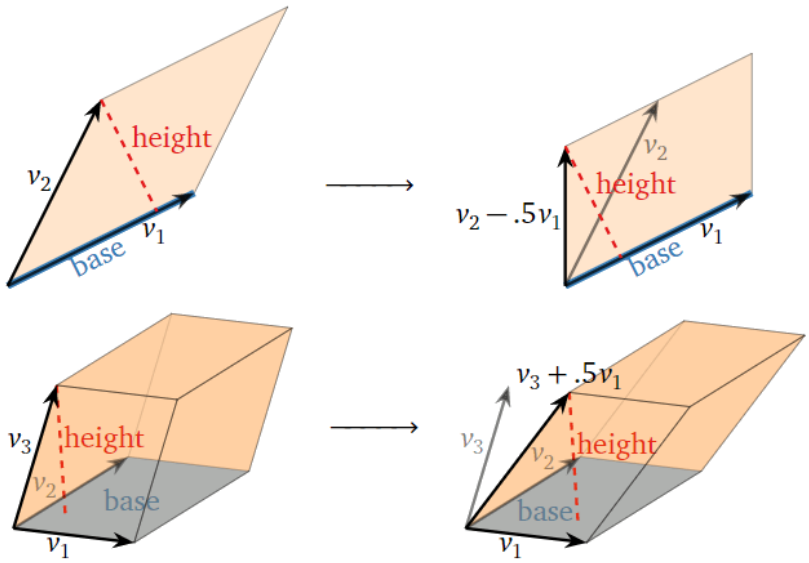


Figure 3.4.7: A diagram demonstrating the invariance of area and volume under row operations. In the top row, two parallelograms are shown. On the left, a parallelogram is spanned by vectors $\vec{v}_1$ and $\vec{v}_2$, where $\vec{v}_1$ is labeled as the "base" in blue, and a red dashed line represents the "height" perpendicular to it. On the right, the parallelogram undergoes a row operation where $\vec{v}_2$ is replaced by $\vec{v}_2 - 0.5\vec{v}_1$, but the height remains unchanged, preserving the area. In the bottom row, two parallelepipeds are shown. On the left, the parallelepiped is defined by vectors $\vec{v}_1, \vec{v}_2$, and $\vec{v}_3$, where $\vec{v}_1$ and $\vec{v}_2$ form the "base" (shaded in gray), and a red dashed line represents the "height" along $\vec{v}_3$. On the right, after the row operation $\vec{v}_3 \to \vec{v}_3 + 0.5\vec{v}_1$, the height remains the same, illustrating that the volume is preserved. (CC-BY-SA; Margalit and Rabinoff via Interactive Linear Algebra)

As we can see, the effect of performing elementary row operations on I_n is identical on the volume of the parallelepiped and the absolute value of the determinant. Hence,

$$|\det(A)| = \text{vol}(P).$$

This page titled 3.4: Determinants and Geometry is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson via source content that was edited to the style and standards of the LibreTexts platform.

- 4.3: Determinants and Volumes by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

3.4E: Exercises for Section 3.4

Exercise 3.4E.1

Find the area of the parallelogram formed by the vectors

$\vec{v}_1 = \langle 3, 1\rangle, \quad \vec{v}_2 = \langle 2, -2\rangle.$

Answer available at the end of this section.

Exercise 3.4E.2

For what values of k do the vectors $\vec{v}_1 = \langle 2, -3\rangle, \quad \vec{v}_2 = \langle -4, k\rangle$ form a parallelogram with zero area?

Answer available at the end of this section.

Exercise 3.4E.3

Find the volume of the parallelepiped formed by

$$\vec{v}_1 = \langle 1, 2, 3\rangle, \quad \vec{v}_2 = \langle 2, 3, 1\rangle, \quad \vec{v}_3 = \langle 3, 1, 2\rangle. \tag{3.4E.1}$$

Answer available at the end of this section.

Exercise 3.4E.4

For what values of k do the vectors

$$\vec{v}_1 = \langle 1, 2, 3\rangle, \quad \vec{v}_2 = \langle 4, k, 6\rangle, \quad \vec{v}_3 = \langle 7, 8, 9\rangle \tag{3.4E.2}$$

form a parallelepiped with zero volume?

Answer available at the end of this section.

Exercise 3.4E.5

Find the area of the triangle with vertices $A(0, 0)$, $B(2, 3)$, and $C(4, 1)$.

Answer available at the end of this section.

Exercise 3.4E.6

Find the area of the triangle with vertices $(1, 2)$, $(4, 6)$, and $(5, 3)$.

Exercise 3.4E.7

For what values of k do the vectors

$$(1, 2, 3), \quad (2, 4, 6), \quad (3, 6, k) \tag{3.4E.3}$$

do not determine a 3 dimensional parallelepiped?

Exercise 3.4E.8

Suppose vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ are the rows of a skew symmetric matrix A. Prove that the volume of the n-dimensional parallelepiped formed by $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ has zero volume.

Hint available at the end of this section.

Exercise 3.4E.9

a. Show that if three vectors in $\mathbb{R}^3$ lie in the same plane, the determinant of their matrix representation is zero.
b. Why does this happen geometrically?
c. Provide an example of three linearly dependent vectors and compute their determinant.

Exercise 3.4E.10

a. Given three points $A(x_1, y_1), B(x_2, y_2), C(x_3, y_3)$, show that they are collinear if and only if

$$\det \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} = 0. \tag{3.4E.4}$$

b. Interpret this result in terms of area.
c. What happens when four points in $\mathbb{R}^3$ satisfy a similar determinant condition?

Exercise 3.4E.11

Suppose we start with the identity matrix I_3. Perform the following operations to obtain matrix A:

- $2R_1 + R_2 -> R_2$
- $3R_2 + R_3 -> R_3$
- $R_2 <-> R_3$
- $-4R_2 + R_1 -> R_1$

a. What is A?
b. What is $\det(A)$?

Answers

Exercise $3.4E.1$
Area = 8

Exercise $3.4E.2$
$k = 6$

Exercise $3.4E.3$
Volume = 18

Exercise $3.4E.4$
$k = 6$.

Exercise $3.4E.5$
Area = 5.

Exercise $3.4E.8$
Since A is skew symmetric, $A^T = -A$.

This page titled 3.4E: Exercises for Section 3.4 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

- 3.4E: Exercises for Section 3.4 by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson is licensed CC BY 4.0.

4: VECTOR SPACES - R^n

4.1: Subspaces

4.2: Review of Vectors

4.2E: Exercises for Section 4.1

4.3: Dot and Cross Product

4.3E: Exercises for Section 4.2

4.4: Lines and Planes

4.4E: Exercises for Section 4.3

4.5: Spanning Sets in R^n

4.5.E: Exercise for Section 4.4

4.6: Linear Independence

4.6.E: Exercise for Section 4.5

4.7: Subspaces

4.8: Subspaces and Bases

4.8.E: Exercise for Section 4.6

4.9: Row, Column and Null Spaces

4.9.E: Exercise for Section 4.7

4.10: Dot Products and Orthogonality

4.11: Orthogonal Vectors and Matrices

4.11.E: Exercise for Section 4.8

4.12: Gram-Schmidt Process

4.12.E: Exercises for Section 4.9

4.13: Orthogonal Complements

4.14: Orthogonal Projections

4.14.E: Exercise for Section 4.10

4.15: Orthogonal Projection

4.16: Least Squares Approximation

4.16.E: Exercises for Section 4.11

4.17: Vector Spaces and Subspace

4.18: Amira Vector Space Axioms

4.19: Amira Vector Subspace of R^n

4.20: The Real Vector Space ^R^n

4.E: Exercises

This page titled 4: Vector Spaces - R^n is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Fatemeh Yarahmadi, De Anza College.

- 4: R^n by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.1: 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.

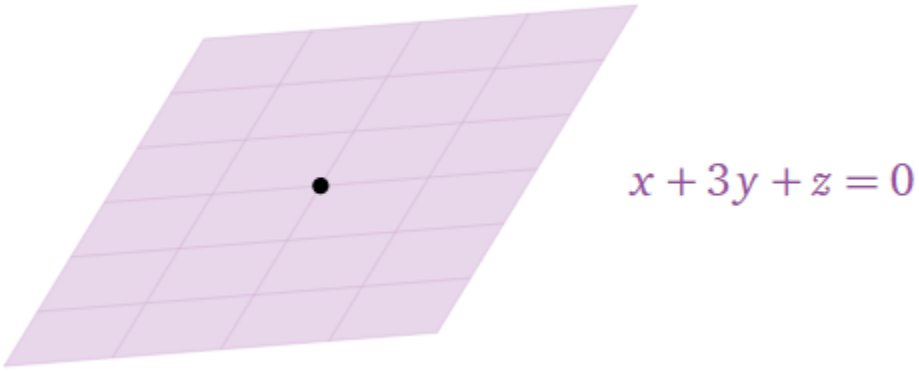


Figure 4.1.1

(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 4.1.1: Subset

A ***subset*** of $\mathbb{R}^n$ is any collection of points of $\mathbb{R}^n$.

For instance, the unit circle

$$C = \{(x, y) \text{ in } \mathbb{R}^2 \mid x^2 + y^2 = 1\}$$

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 4.1.2: 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 4.1.1 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 4.1.1

The set $\mathbb{R}^n$ is a subspace of itself: indeed, it contains zero, and is closed under addition and scalar multiplication.

Example 4.1.2

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 4.1.3: A line through the origin

A line L through the origin is a subspace.

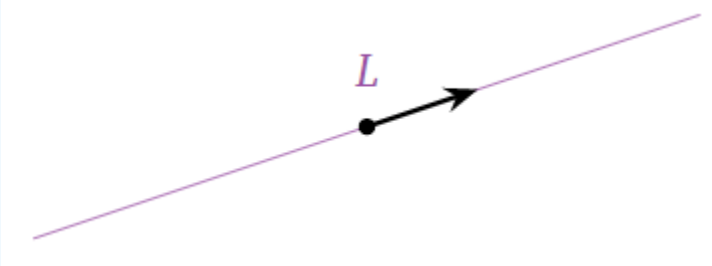


Figure 4.1.2

Indeed, L contains zero, and is easily seen to be closed under addition and scalar multiplication.

Example 4.1.4: A plane through the origin

A plane P through the origin is a subspace.

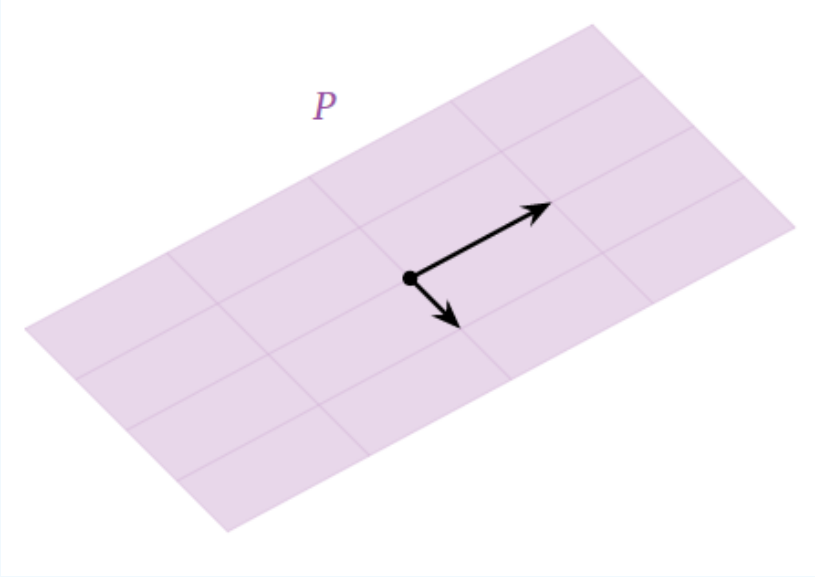


Figure 4.1.3

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 4.1.5: 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.

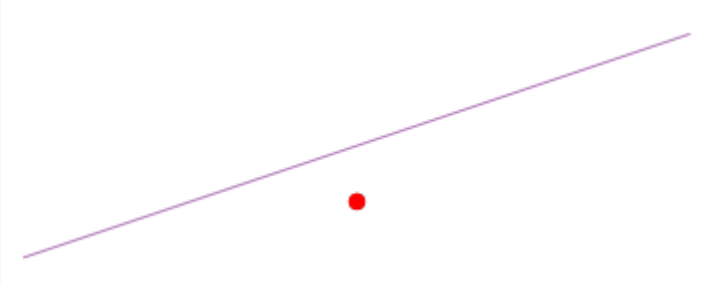


Figure 4.1.4

Example 4.1.6: 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.

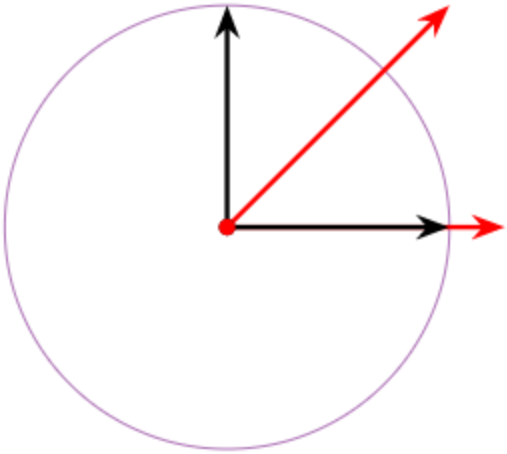


Figure 4.1.5

Example 4.1.7: 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).

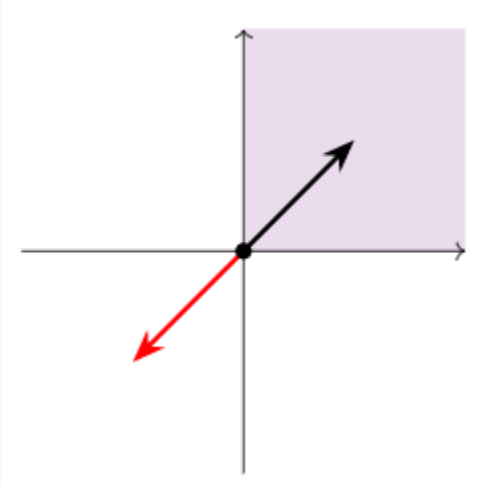


Figure 4.1.6

Example 4.1.8: 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.

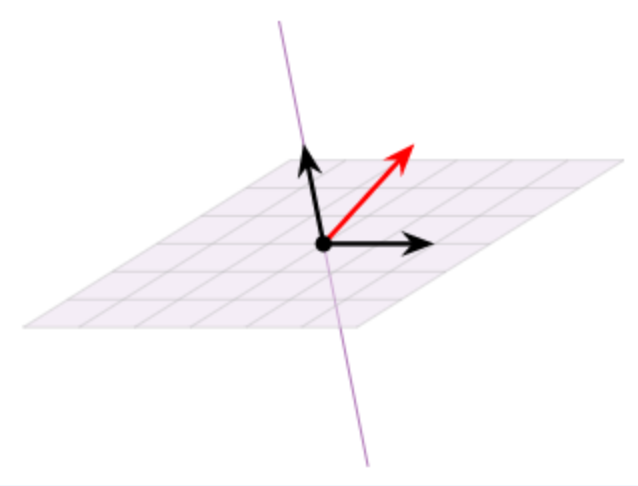


Figure 4.1.7

Note 4.1.1: Subsets versus Subspaces

A subset of $\mathbb{R}^n$ is any collection of vectors whatsoever. For instance, the unit circle

$$C = \left\{(x, y) \text{ in } \mathbb{R}^2 \mid x^2 + y^2 = 1\right\}$$

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 4.1.3, below.

Example 4.1.9: 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 $\begin{pmatrix} a \\ b \end{pmatrix}$ 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, $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 1/2 \\ 1/3 \end{pmatrix}$ are in V, but $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ is not because $2 \cdot 2 \neq 3 \cdot 3$.

Let us check the first property. The subset V does contain the zero vector $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, 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 $u = \begin{pmatrix} a \\ b \end{pmatrix}$ and $v = \begin{pmatrix} c \\ d \end{pmatrix}$ 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 $\begin{pmatrix} a+c \\ b+d \end{pmatrix}$ 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 $v = \begin{pmatrix} a \\ b \end{pmatrix}$ 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 $\begin{pmatrix} ca \\ cb \end{pmatrix}$ 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 $2/3$.

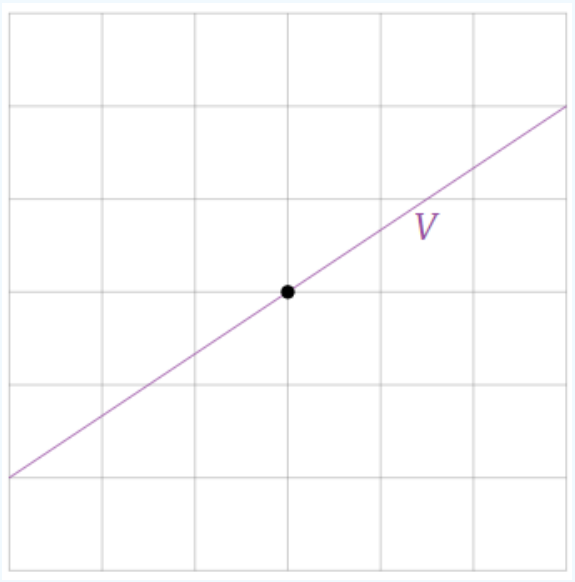


Figure 4.1.8

Example 4.1.10: 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.

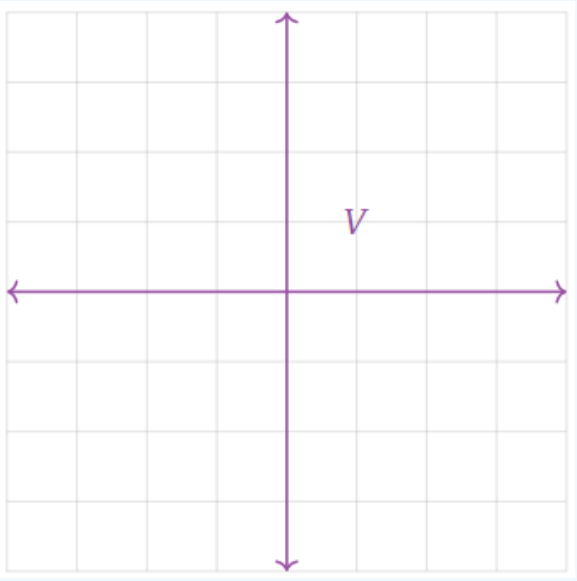


Figure 4.1.9

Common Types of Subspaces

Theorem 4.1.1: 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 available at the end of this section.

If $V = \text{Span}\{v_1, v_2, \dots, v_p\}$, we say that V is the subspace ***spanned by*** or ***generated by*** the vectors $v_1, v_2, \dots, v_p$. We call $\{v_1, v_2, \dots, v_p\}$ a ***spanning set*** for V.

Any matrix naturally gives rise to *two* subspaces.

Definition 4.1.3: 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 4.1.1. 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 4.1.11

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 $\mathbb{R}^3$.

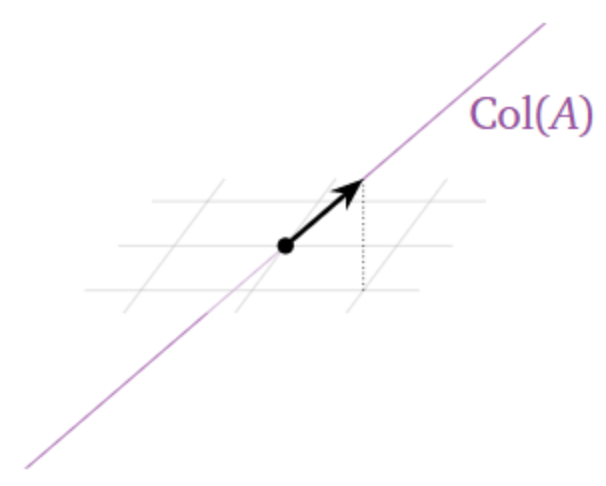


Figure 4.1.10

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 $\mathbb{R}^2$.

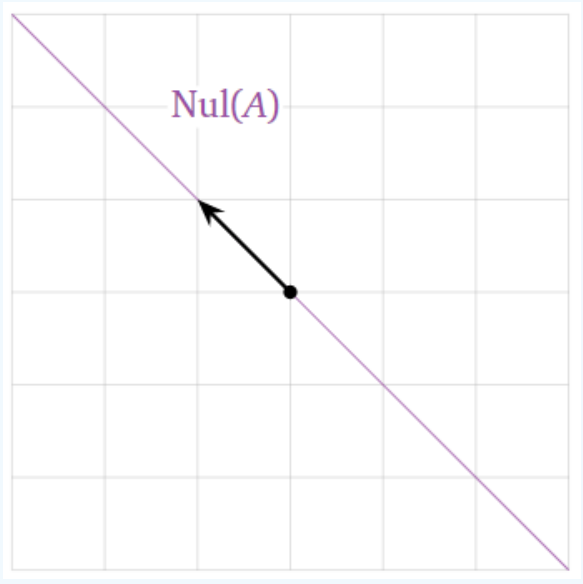


Figure 4.1.11

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 4.1.2: 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 4.1.12: 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,

$$\text{Nul}(A) = \text{Span}\left\{\begin{pmatrix}1\\2\\1\\0\end{pmatrix}, \begin{pmatrix}-2\\3\\0\\1\end{pmatrix}\right\}.$$

Example 4.1.13: 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,

$$\text{Nul}(A) = \{0\} = \text{Span}\{0\}.$$

It is natural to define $\text{Span}\{\} = \{0\}$, 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 4.1.3

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 4.1.14

Let

$$V = \left\{\begin{pmatrix}a\\b\end{pmatrix} \text{ in } \mathbb{R}^2 | 2a = 3b\right\}$$

be the subset of a previous Example 4.1.9. 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 $\begin{pmatrix} 2 & -3 \end{pmatrix}$ is $\begin{pmatrix} 1 & -3/2 \end{pmatrix}$, so the parametric form of V is $x = 3/2y$, so the parametric vector form is $\begin{pmatrix} x \\ y \end{pmatrix} = y\begin{pmatrix} 3/2 \\ 1 \end{pmatrix}$, and hence $\left\{\begin{pmatrix} 3/2 \\ 1 \end{pmatrix}\right\}$ spans V.

Example 4.1.15

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

$$V = \text{Col}\begin{pmatrix} 2 & 1 \\ 1 & -1 \\ 3 & -2 \end{pmatrix}.$$

Answers

Theorem 4.1.1: Spans are Subspaces and Subspaces are Spans

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.

This page titled 4.1: Subspaces is shared under a GNU FDL license and was authored, remixed, and/or curated by Dan Margalit and Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- 2.6: Subspaces by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

4.2: Review of Vectors

Learning Objectives

- Find the position vector of a point in $\mathbb{R}^n$.
- Understand vector addition and scalar multiplication, algebraically.
- Understand vector addition, geometrically.
- Find the length of a vector and the distance between two points in $\mathbb{R}^n$.
- Find the corresponding unit vector to a vector in $\mathbb{R}^n$.
- Understand scalar multiplication, geometrically.
- Introduce the notion of linear combination of vectors.

In this section we review the concept and operations between vectors.

The notation $\mathbb{R}^n$ refers to the collection of ordered lists of n real numbers, that is

$$\mathbb{R}^n = \left\{(x_1, \cdots, x_n) : x_j^{\text{int}}\mathbb{R} \text{ for } j = 1, \cdots, n\right\}$$

In this chapter, we take a closer look at vectors in $\mathbb{R}^n$. First, we will consider what $\mathbb{R}^n$ looks like in more detail. Recall that the point given by $0 = (0, \cdots, 0)$ is called the **origin**.

Now, consider the case of $\mathbb{R}^n$ for $n = 1$. Then from the definition we can identify $\mathbb{R}$ with points in $\mathbb{R}^1$ as follows:

$$\mathbb{R} = \mathbb{R}^1 = \left\{(x_1) : x_1^{\text{int}}\mathbb{R}\right\}$$

Hence, $\mathbb{R}$ is defined as the set of all real numbers and geometrically, we can describe this as all the points on a line.

Now suppose $n = 2$. Then, from the definition,

$$\mathbb{R}^2 = \left\{(x_1, x_2) : x_j^{\text{int}}\mathbb{R} \text{ for } j = 1, 2\right\}$$

Consider the familiar coordinate plane, with an x axis and a y axis. Any point within this coordinate plane is identified by where it is located along the x axis, and also where it is located along the y axis. Consider as an example the following diagram.

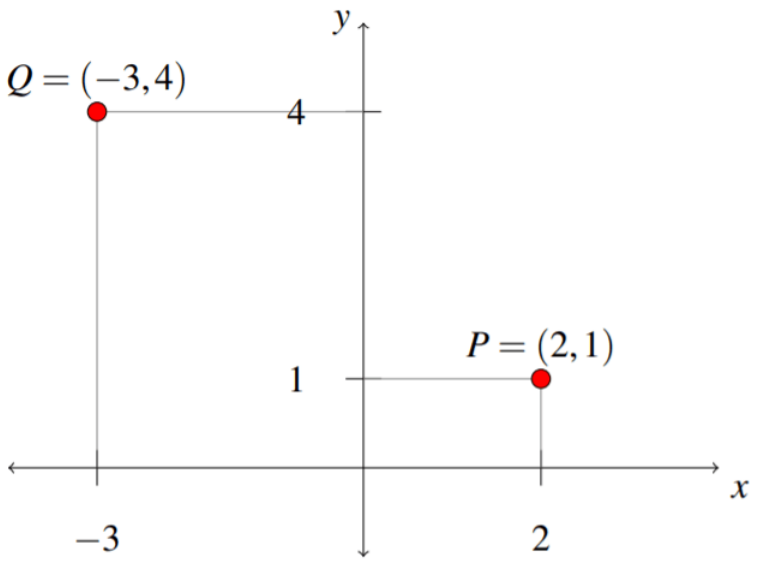


Figure 4.1.1: Diagram illustrates points $Q = (-3, 4)$ and $P = (2, 1)$ in $\mathbb{R}^2$. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Hence, every element in $\mathbb{R}^2$ is identified by two components, x and y, in the usual manner. The coordinates x, y (or x_1,x_2) uniquely determine a point in the plan. Note that while the definition uses x_1 and x_2 to label the coordinates and you may be used to x and y, these notations are equivalent.

Now suppose $n = 3$. You may have previously encountered the 3-dimensional coordinate system, given by

$$\mathbb{R}^3 = \left\{(x_1, x_2, x_3) : x_j \overset{\text{int}}{\mathbb{R}} \text{ for } j = 1, 2, 3\right\}$$

Points in $\mathbb{R}^3$ will be determined by three coordinates, often written (x, y, z) which correspond to the x, y, and z axes. We can think as above that the first two coordinates determine a point in a plane. The third component determines the height above or below the plane, depending on whether this number is positive or negative, and all together this determines a point in space. You see that the ordered triples correspond to points in space just as the ordered pairs correspond to points in a plane and single real numbers correspond to points on a line.

The idea behind the more general $\mathbb{R}^n$ is that we can extend these ideas beyond $n = 3$. This discussion regarding points in $\mathbb{R}^n$ leads into a study of vectors in $\mathbb{R}^n$. While we consider $\mathbb{R}^n$ for all n, we will largely focus on $n = 2, 3$ in this section.

Consider the following definition.

Definition 4.2.1 The Position Vector

Let $P = (p_1, \cdots, p_n)$ be the coordinates of a point in $\mathbb{R}^n$. Then the vector $\overrightarrow{0P}$ with its tail at $0 = (0, \cdots, 0)$ and its tip at P is called the **position vector** of the point P. We write

$$\overrightarrow{0P} = \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix}$$

For this reason we may write both $P = (p_1, \cdots, p_n) \overset{\text{int}}{\mathbb{R}^n}$ and Double exponent: use braces to clarify .

This definition is illustrated in the following picture for the special case of $\mathbb{R}^3$.

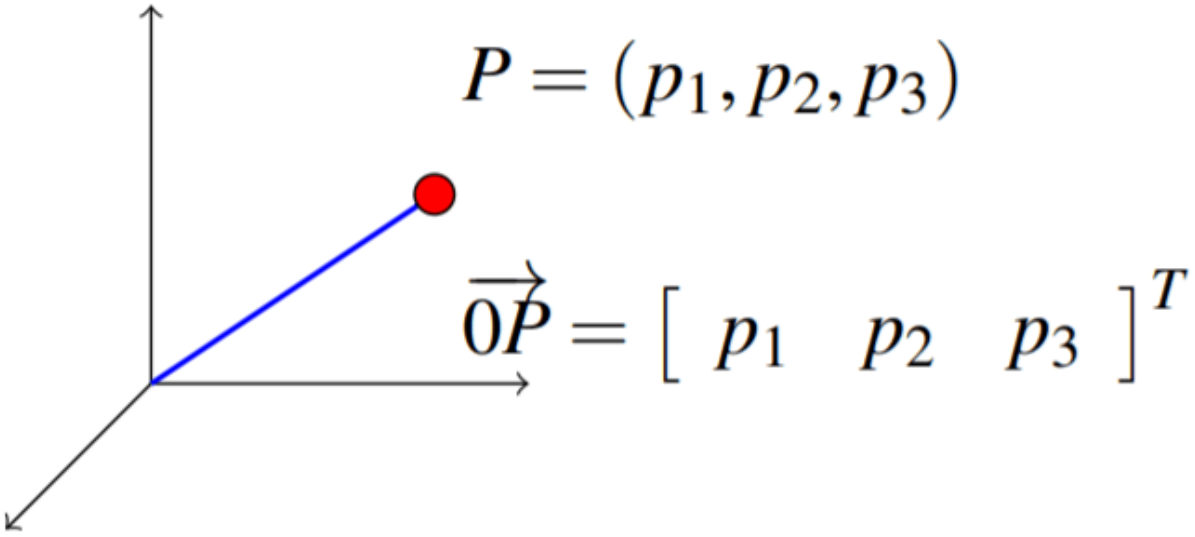


Figure 4.1.2: Diagram illustrates a 3D coordinate system where $P(p_1, p_2, p_3)$ is a point in space. The position vector $\vec{OP}$ extends from the origin O to the point P. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Thus every point P in $\mathbb{R}^n$ determines its position vector $\overrightarrow{0P}$. Conversely, every such position vector $\overrightarrow{0P}$ which has its tail at 0 and point at P determines the point P of $\mathbb{R}^n$.

Now suppose we are given two points, P, Q whose coordinates are $(p_1, \cdots, p_n)$ and $(q_1, \cdots, q_n)$ respectively. We can also determine the **position vector from** P **to** Q (also called the **vector from** P **to** Q) defined as follows.

$$\overrightarrow{PQ} = \begin{bmatrix} q_1 - p_1 \\ \vdots \\ q_n - p_n \end{bmatrix} = \overrightarrow{0Q} - \overrightarrow{0P}$$

Now, imagine taking a vector in $\mathbb{R}^n$ and moving it around, always keeping it pointing in the same direction as shown in the following picture.

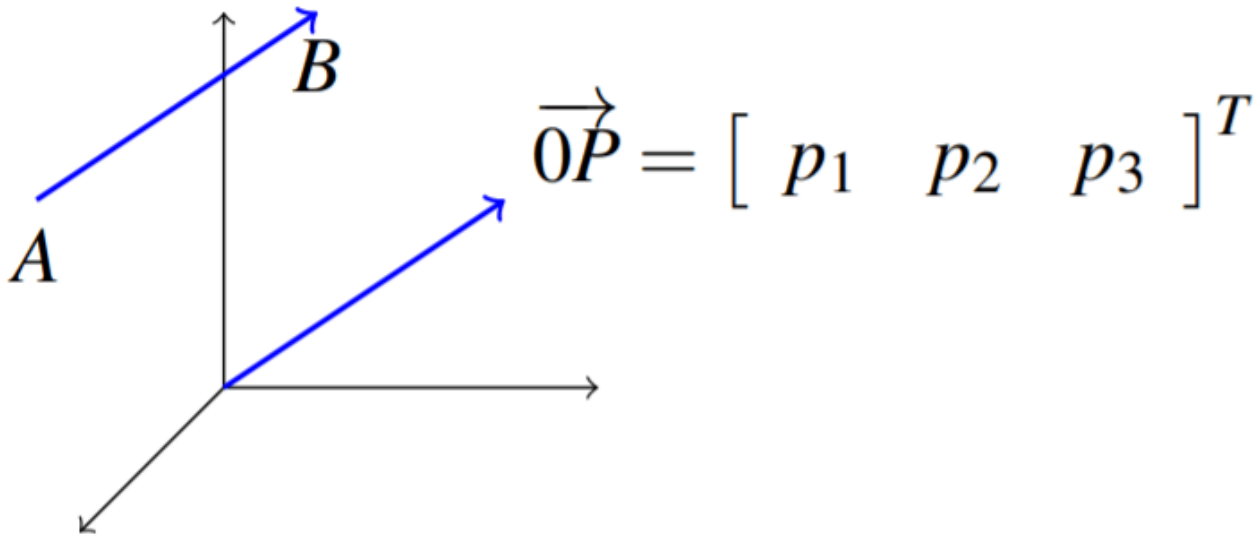


Figure 4.1.3: Diagram illustrates vectors in $\mathbb{R}^3$. $\vec{OP}$ is the position vector from O to point P. $\vec{AB}$ is a vector parallel to $\vec{OP}$. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

After moving it around, it is regarded as the same vector. Each vector, $\overrightarrow{0P}$ and $\overrightarrow{AB}$ has the same length (or magnitude) and direction. Therefore, they are equal.

Consider now the general definition for a vector in $\mathbb{R}^n$.

Definition 4.2.2 Vectors in $\mathbb{R}^n$

Let $\mathbb{R}^n = \left\{(x_1, \cdots, x_n) : x_j^{\text{int}}\mathbb{R} \text{ for } j = 1, \cdots, n\right\}$. Then,

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

is called a **vector**. Vectors have both size (magnitude) and direction. The numbers x_j are called the **components** of $\vec{x}$.

Using this notation, we may use $\vec{p}$ to denote the position vector of point P. Notice that in this context, $\vec{p} = \overrightarrow{0P}$. These notations may be used interchangeably.

You can think of the components of a vector as directions for obtaining the vector. Consider $n = 3$. Draw a vector with its tail at the point $(0, 0, 0)$ and its tip at the point (a, b, c). This vector it is obtained by starting at $(0, 0, 0)$, moving parallel to the x axis to $(a, 0, 0)$ and then from here, moving parallel to the y axis to $(a, b, 0)$ and finally parallel to the z axis to (a, b, c). Observe that the same vector would result if you began at the point (d, e, f), moved parallel to the x axis to $(d + a, e, f)$, then parallel to the y axis to $(d + a, e + b, f)$, and finally parallel to the z axis to $(d + a, e + b, f + c)$. Here, the vector would have its tail sitting at the point determined by $A = (d, e, f)$ and its point at $B = (d + a, e + b, f + c)$. It is the **same vector** because it will point in the same direction and have the same length. It is like you took an actual arrow, and moved it from one location to another keeping it pointing the same direction.

We conclude this section with a brief discussion regarding notation. In previous sections, we have written vectors as columns, or

$n \times 1$ matrices. For convenience in this chapter we may write vectors as the transpose of row vectors, or $1 \times n$ matrices. These are of course equivalent and we may move between both notations. Therefore, recognize that

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

Notice that two vectors $\vec{u} = [u_1 \cdots u_n]^T$ and $\vec{v} = [v_1 \cdots v_n]^T$ are equal if and only if all corresponding components are equal. Precisely,

$$\begin{aligned} &\vec{u} = \vec{v} \text{ if and only if} \\ &u_j = v_j \text{ for all } j = 1, \cdots, n \end{aligned}$$

Thus Double exponent: use braces to clarify and Double exponent: use braces to clarify but $\begin{bmatrix} 1 & 2 & 4 \end{bmatrix}^T \neq \begin{bmatrix} 2 & 1 & 4 \end{bmatrix}^T$ because, even though the same numbers are involved, the order of the numbers is different.

For the specific case of $\mathbb{R}^3$, there are three special vectors which we often use. They are given by

$$\begin{aligned} \vec{i} &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T \\ \vec{j} &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T \\ \vec{k} &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T \end{aligned}$$

We can write any vector $\vec{u} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}^T$ as a linear combination of these vectors, written as $\vec{u} = u_1\vec{i} + u_2\vec{j} + u_3\vec{k}$. This notation will be used throughout this chapter.

Addition and scalar multiplication are two important algebraic operations done with vectors. Notice that these operations apply to vectors in $\mathbb{R}^n$, for any value of n. We will explore these operations in more detail in the following sections.

Recall that an element of $\mathbb{R}^n$ is an ordered list of numbers. For the specific case of $n = 2, 3$ this can be used to determine a point in two or three dimensional space. This point is specified relative to some coordinate axes.

Consider the case $n = 3$. Recall that taking a vector and moving it around without changing its length or direction does not change the vector. This is important in the geometric representation of vector addition.

Suppose we have two vectors, $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^3$. Each of these can be drawn geometrically by placing the tail of each vector at 0 and its point at (u_1, u_2, u_3) and (v_1, v_2, v_3) respectively. Suppose we slide the vector $\vec{v}$ so that its tail sits at the point of $\vec{u}$. We know that this does not change the vector $\vec{v}$. Now, draw a new vector from the tail of $\vec{u}$ to the point of $\vec{v}$. This vector is $\vec{u} + \vec{v}$.

The geometric significance of vector addition in $\mathbb{R}^n$ for any n is given in the following definition.

Definition 4.2.3: Geometry of Vector Addition

Let $\vec{u}$ and $\vec{v}$ be two vectors. Slide $\vec{v}$ so that the tail of $\vec{v}$ is on the point of $\vec{u}$. Then draw the arrow which goes from the tail of $\vec{u}$ to the point of $\vec{v}$. This arrow represents the vector $\vec{u} + \vec{v}$.

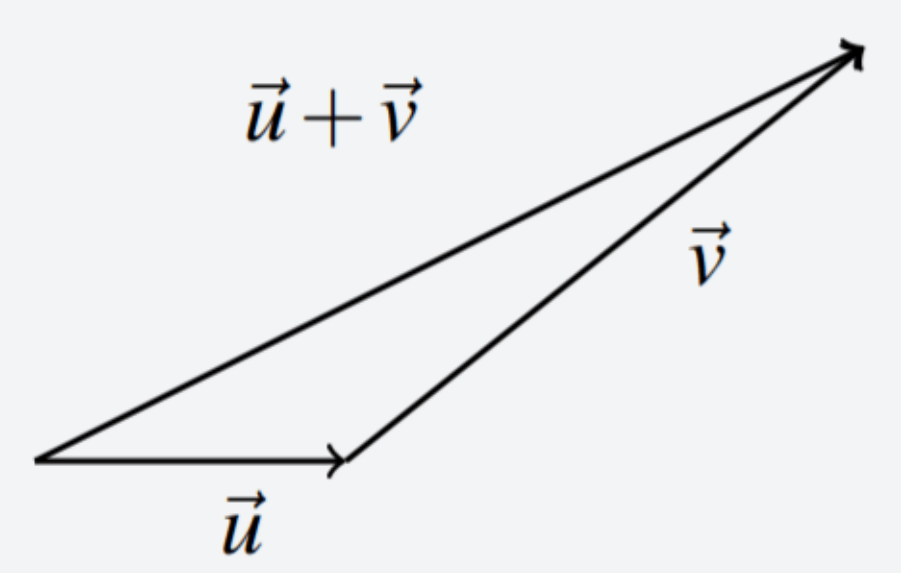


Figure 4.2.4: Diagram illustrates vector addition in $\mathbb{R}^2$. The vector $\vec{u}$ starts at the origin, the vector $\vec{v}$ is placed at the tip of $\vec{u}$, and their sum $\vec{u}$ + $\vec{v}$ is the resulting vector from the origin to the tip of $\vec{v}$. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

This definition is illustrated in the following picture in which $\vec{u}+\vec{v}$ is shown for the special case $n=3$.

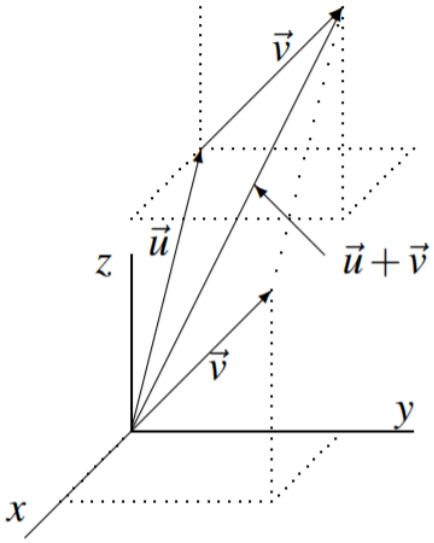


Figure 4.2.5: Diagram illustrating vector addition in $\mathbb{R}^3$, showing two vectors $\vec{u}$ and $\vec{v}$ in three-dimensional space. The image demonstrates how to add two vectors geometrically by placing the tail of one vector at the tip of the other. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Notice the parallelogram created by $\vec{u}$ and $\vec{v}$ in the above diagram. Then $\vec{u}+\vec{v}$ is the directed diagonal of the parallelogram determined by the two vectors $\vec{u}$ and $\vec{v}$.

When you have a vector $\vec{v}$, its additive inverse $-\vec{v}$ will be the vector which has the same magnitude as $\vec{v}$ but the opposite direction. When one writes $\vec{u}-\vec{v}$, the meaning is $\vec{u}+(-\vec{v})$ as with real numbers. The following example illustrates these definitions and conventions.

Example 4.2.1: Graphing Vector Addition

Consider the following picture of vectors $\vec{u}$ and $\vec{v}$.



Figure 4.2.6: Diagram illustrates two vectors in $\mathbb{R}^2$. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Sketch a picture of $\vec{u}+\vec{v}, \vec{u}-\vec{v}$.

Solution

We will first sketch $\vec{u}+\vec{v}$. Begin by drawing $\vec{u}$ and then at the point of $\vec{u}$, place the tail of $\vec{v}$ as shown. Then $\vec{u}+\vec{v}$ is the vector which results from drawing a vector from the tail of $\vec{u}$ to the tip of $\vec{v}$.

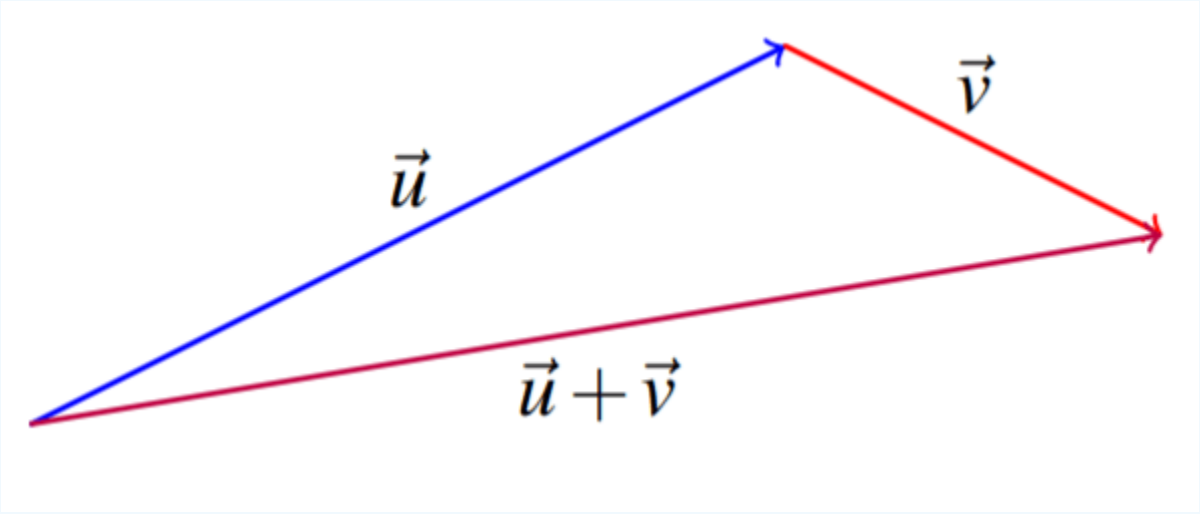


Figure 4.2.7: Diagram illustrates vector addition in $\mathbb{R}^2$, showing two vectors $\vec{u}$ and $\vec{v}$. The image demonstrates how to add two vectors geometrically by placing the tail of one vector at the tip of the other. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Next consider $\vec{u}-\vec{v}$. This means $\vec{u}+\left(-\vec{v}\right)$. From the above geometric description of vector addition, $-\vec{v}$ is the vector which has the same length but which points in the opposite direction to $\vec{v}$. Here is a picture.

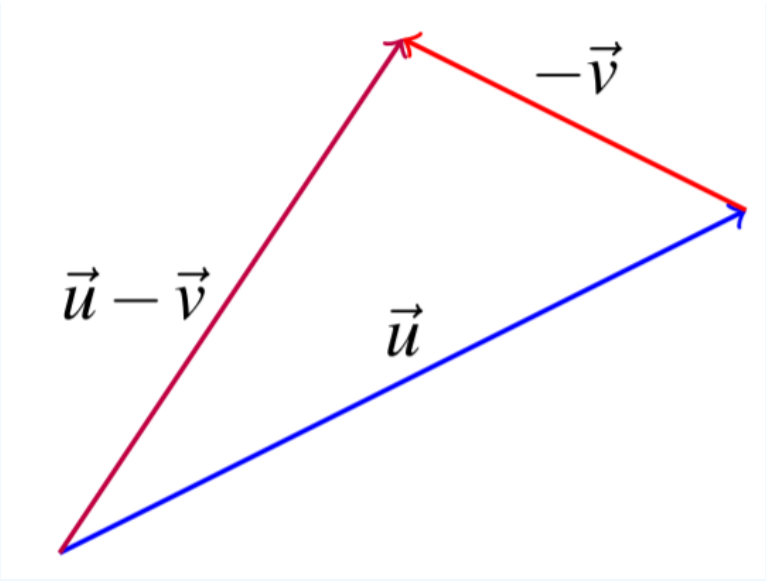


Figure 4.2.8: Diagram illustrates vector subtraction in $\mathbb{R}^2$, showing two vectors $\vec{u}$ and $\vec{v}$. The image demonstrates how to subtract two vectors geometrically by placing the tail of $-\vec{v}$ at the tip of $\vec{u}$. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Addition of Vectors in $\mathbb{R}^n$

Addition of vectors in $\mathbb{R}^n$ is defined as follows.

Definition 4.2.4: Addition of Vectors in $\mathbb{R}^n$

If $\vec{u}=\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$, $\vec{v}=\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ $\in^{\text{int}} \mathbb{R}^n$ then $\vec{u}+\vec{v} \in^{\text{int}} \mathbb{R}^n$ and is defined by

$$\begin{aligned} \vec{u}+\vec{v} &= \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \\ &= \begin{bmatrix} u_1+v_1 \\ \vdots \\ u_n+v_n \end{bmatrix} \end{aligned} \tag{4.2.1}$$

To add vectors, we simply add corresponding components. Therefore, in order to add vectors, they must be the same size.

Addition of vectors satisfies some important properties which are outlined in the following theorem.

Theorem 4.2.1: Properties of Vector Addition

The following properties hold for vectors $\vec{u},\vec{v},\vec{w} \in^{\text{int}} \mathbb{R}^n$.

- The Commutative Law of Addition

$$\vec{u}+\vec{v}=\vec{v}+\vec{u}$$

- The Associative Law of Addition

$$(\vec{u}+\vec{v})+\vec{w}=\vec{u}+(\vec{v}+\vec{w})$$

- The Existence of an Additive Identity

$$\vec{u} + \vec{0} = \vec{u} \tag{4.2.2}$$

- The Existence of an Additive Inverse

$$\vec{u} + (-\vec{u}) = \vec{0}$$

The additive identity shown in Equation (4.2.2) is also called the **zero vector**, the $n \times 1$ vector in which all components are equal to 0. Further, $-\vec{u}$ is simply the vector with all components having same value as those of $\vec{u}$ but opposite sign; this is just $(-1)\vec{u}$. This will be made more explicit in the next section when we explore scalar multiplication of vectors. Note that subtraction is defined as $\vec{u} - \vec{v} = \vec{u} + (-\vec{v})$.

We now explore what is meant by the length of a vector in $\mathbb{R}^n$. We develop this concept by first looking at the distance between two points in $\mathbb{R}^n$.

First, we will consider the concept of distance for $\mathbb{R}$, that is, for points in $\mathbb{R}^1$. Here, the distance between two points P and Q is given by the absolute value of their difference. We denote the distance between P and Q by $d(P,Q)$ which is defined as

$$d(P,Q) = \sqrt{(P-Q)^2} \tag{4.2.3}$$

Consider now the case for $n = 2$, demonstrated by the following picture.

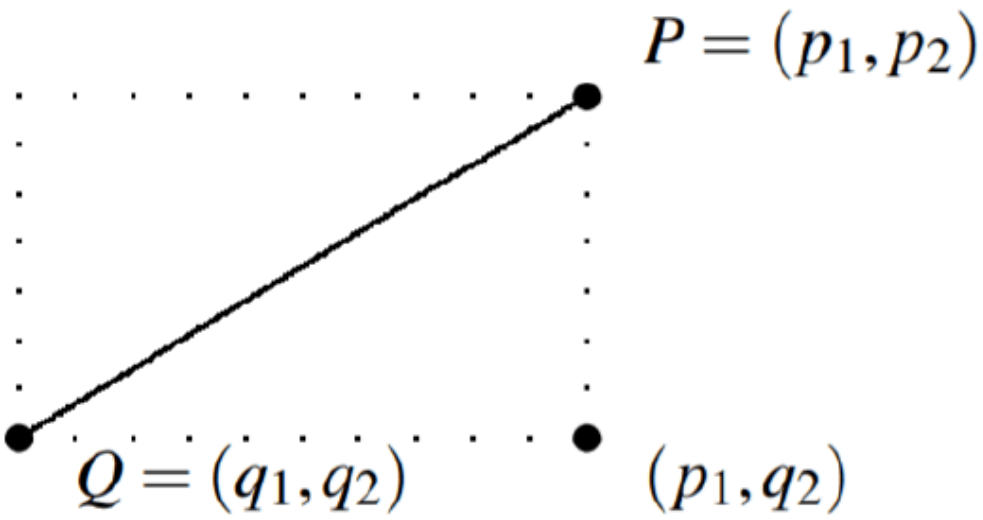


Figure 4.2.9: Diagram illustrates the distance between two points P and Q in $\mathbb{R}^2$. The image demonstrates how to calculate the distance using the Pythagorean theorem. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

There are two points $P = (p_1, p_2)$ and $Q = (q_1, q_2)$ in the plane. The distance between these points is shown in the picture as a solid line. Notice that this line is the hypotenuse of a right triangle which is half of the rectangle shown in dotted lines. We want to find the length of this hypotenuse which will give the distance between the two points. Note the lengths of the sides of this triangle are $|p_1 - q_1|$ and $|p_2 - q_2|$, the absolute value of the difference in these values. Therefore, the Pythagorean Theorem implies the length of the hypotenuse (and thus the distance between P and Q) equals

$$\left(|p_1 - q_1|^2 + |p_2 - q_2|^2\right)^{1/2} = \left((p_1 - q_1)^2 + (p_2 - q_2)^2\right)^{1/2} \tag{4.2.4}$$

Now suppose $n = 3$ and let $P = (p_1, p_2, p_3)$ and $Q = (q_1, q_2, q_3)$ be two points in $\mathbb{R}^3$. Consider the following picture in which the solid line joins the two points and a dotted line joins the points (q_1, q_2, q_3) and (p_1, p_2, q_3) .

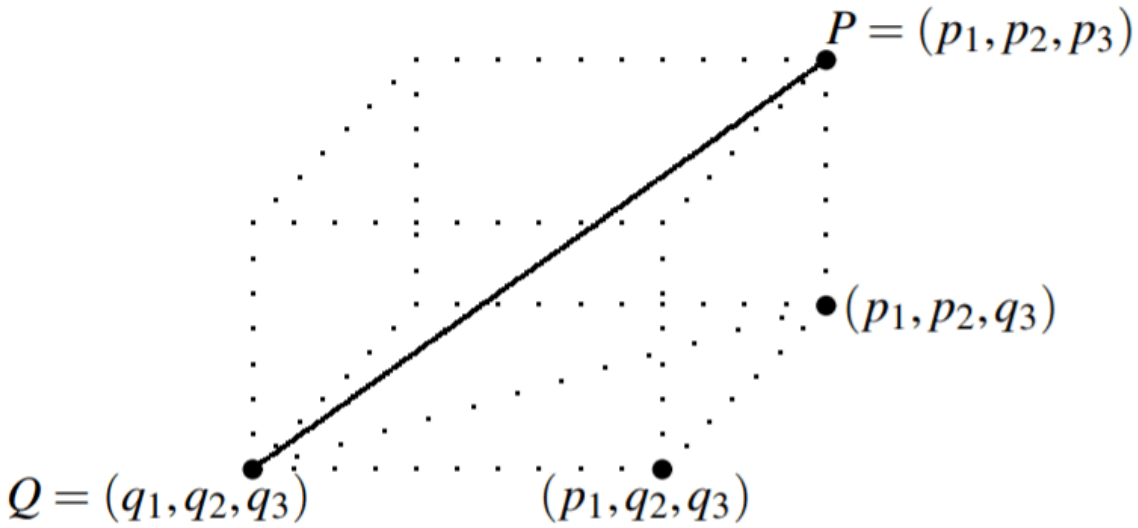


Figure 4.2.10: Diagram illustrates the distance between two points P and Q in $\mathbb{R}^3$. The image demonstrates how to calculate the distance using the Pythagorean theorem in three-dimensional space. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Here, we need to use Pythagorean Theorem twice in order to find the length of the solid line. First, by the Pythagorean Theorem, the length of the dotted line joining (q_1, q_2, q_3) and (p_1, p_2, q_3) equals

$$\left((p_1 - q_1)^2 + (p_2 - q_2)^2\right)^{1/2}$$

while the length of the line joining (p_1, p_2, q_3) to (p_1, p_2, p_3) is just $|p_3 - q_3|$. Therefore, by the Pythagorean Theorem again, the length of the line joining the points $P = (p_1, p_2, p_3)$ and $Q = (q_1, q_2, q_3)$ equals

$$\begin{aligned} &\left(\left(\left((p_1 - q_1)^2 + (p_2 - q_2)^2\right)^{1/2}\right)^2 + (p_3 - q_3)^2\right)^{1/2} \\ &= \left((p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2\right)^{1/2} \end{aligned} \tag{4.2.5}$$

This discussion motivates the following definition for the distance between points in $\mathbb{R}^n$.

Definition 4.2.5: Distance Between Points

Let $P = (p_1, \cdots, p_n)$ and $Q = (q_1, \cdots, q_n)$ be two points in $\mathbb{R}^n$. Then the distance between these points is defined as

$$\text{distance between } P \text{ and } Q \ = d(P, Q) = \left(\sum_{k=1}^{n} |p_k - q_k|^2\right)^{1/2}$$

This is called the **distance formula**. We may also write $|P - Q|$ as the distance between P and Q.

From the above discussion, you can see that Definition 4.2.1 holds for the special cases $n = 1, 2, 3$, as in Equations (4.2.3), (4.2.4), (4.2.5). In the following example, we use Definition 4.2.1 to find the distance between two points in $\mathbb{R}^4$.

Example 4.2.2: Distance Between Points

Find the distance between the points P and Q in $\mathbb{R}^4$, where P and Q are given by

$$P = (1, 2, -4, 6)$$

and

$$Q = (2, 3, -1, 0)$$

Solution

We will use the formula given in Definition 4.2.1 to find the distance between P and Q. Use the distance formula and write

$$d(P,Q) = \left((1-2)^2 + (2-3)^2 + (-4-(-1))^2 + (6-0)^2\right)^{\frac{1}{2}} = 47$$

Therefore, $d(P,Q) = \sqrt{47}$.

There are certain properties of the distance between points which are important in our study. These are outlined in the following theorem.

Theorem 4.2.2: Properties of Distance

Let P and Q be points in $\mathbb{R}^n$, and let the distance between them, $d(P,Q)$, be given as in Definition 4.2.1. Then, the following properties hold.

- $d(P,Q) = d(Q,P)$
- $d(P,Q) \geq 0$, and equals 0 exactly when $P = Q$.

There are many applications of the concept of distance. For instance, given two points, we can ask what collection of points are all the same distance between the given points. This is explored in the following example.

Example 4.2.3: The Plane Between Two Points

Describe the points in $\mathbb{R}^3$ which are at the same distance between $(1,2,3)$ and $(0,1,2)$.

Solution

Let $P = (p_1, p_2, p_3)$ be such a point. Therefore, P is the same distance from $(1,2,3)$ and $(0,1,2)$. Then byDefinition 4.2.1,

$$\sqrt{(p_1-1)^2 + (p_2-2)^2 + (p_3-3)^2} = \sqrt{(p_1-0)^2 + (p_2-1)^2 + (p_3-2)^2}$$

Squaring both sides we obtain

$$(p_1-1)^2 + (p_2-2)^2 + (p_3-3)^2 = p_1^2 + (p_2-1)^2 + (p_3-2)^2$$

and so

$$p_1^2 - 2p_1 + 14 + p_2^2 - 4p_2 + p_3^2 - 6p_3 = p_1^2 + p_2^2 - 2p_2 + 5 + p_3^2 - 4p_3$$

Simplifying, this becomes

$$-2p_1 + 14 - 4p_2 - 6p_3 = -2p_2 + 5 - 4p_3$$

which can be written as

$$2p_1 + 2p_2 + 2p_3 = -9 \tag{4.2.6}$$

Therefore, the points $P = (p_1, p_2, p_3)$ which are the same distance from each of the given points form a plane whose equation is given by (4.2.6).

We can now use our understanding of the distance between two points to define what is meant by the length of a vector. Consider the following definition.

Definition 4.2.6: Length of a Vector

Let $\vec{u} = [u_1 \cdots u_n]^T$ be a vector in $\mathbb{R}^n$. Then, the length of $\vec{u}$, written $\|\vec{u}\|$ is given by

$$\|\vec{u}\| = \sqrt{u_1^2 + \cdots + u_n^2}$$

This definition corresponds to Definition 4.2.1, if you consider the vector $\vec{u}$ to have its tail at the point $0 = (0, \cdots, 0)$ and its tip at

the point $U = (u_1, \cdots, u_n)$. Then the length of $\vec{u}$ is equal to the distance between 0 and U, $d(0, U)$. In general, $d(P, Q) = ||\vec{PQ}||$.

ConsiderExample $4.2.1$. ByDefinition $4.2.2$, we could also find the distance between P and Q as the length of the vector connecting them. Hence, if we were to draw a vector $\overrightarrow{PQ}$ with its tail at P and its point at Q, this vector would have length equal to $\sqrt{47}$.

We conclude this section with a new definition for the special case of vectors of length 1.

Definition $4.2.7$: Unit Vector

Let $\vec{u}$ be a vector in $\mathbb{R}^n$. Then, we call $\vec{u}$ a **unit vector** if it has length 1, that is if

$$\|\vec{u}\| = 1$$

Let $\vec{v}$ be a vector in $\mathbb{R}^n$. Then, the vector $\vec{u}$ which has the same direction as $\vec{v}$ but length equal to 1 is the corresponding unit vector of $\vec{v}$. This vector is given by

$$\vec{u} = \frac{1}{\|\vec{v}\|}\vec{v}$$

We often use the term **normalize** to refer to this process. When we **normalize** a vector, we find the corresponding unit vector of length 1. Consider the following example.

Example $4.2.4$: Finding a Unit Vector

Let $\vec{v}$ be given by

$$\vec{v} = \begin{bmatrix} 1 & -3 & 4 \end{bmatrix}^T$$

Find the unit vector $\vec{u}$ which has the same direction as $\vec{v}$.

Solution

We will use Definition $4.2.3$ to solve this. Therefore, we need to find the length of $\vec{v}$ which, by Definition $4.2.2$ is given by

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2}$$

Using the corresponding values we find that

$$\begin{aligned} \|\vec{v}\| &= \sqrt{1^2 + (-3)^2 + 4^2} \\ &= \sqrt{1 + 9 + 16} \\ &= \sqrt{26} \end{aligned} \tag{4.2.7}$$

In order to find $\vec{u}$, we divide $\vec{v}$ by $\sqrt{26}$. The result is

$$\begin{aligned} \vec{u} &= \frac{1}{\|\vec{v}\|}\vec{v} \\ &= \frac{1}{\sqrt{26}}\begin{bmatrix} 1 & -3 & 4 \end{bmatrix}^T \\ &= \begin{bmatrix} \frac{1}{\sqrt{26}} & -\frac{3}{\sqrt{26}} & \frac{4}{\sqrt{26}} \end{bmatrix}^T \end{aligned} \tag{4.2.8}$$

You can verify using the Definition $4.2.1$ that $\|\vec{u}\| = 1$.

Now suppose we have a vector $\vec{u} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}^T$ and we multiply $\vec{u}$ by a scalar k. By Definition 4.2.2, $k\vec{u} = \begin{bmatrix} ku_1 & ku_2 & ku_3 \end{bmatrix}^T$. Then, by using Definition 4.4.1, the length of this vector is given by

$$\sqrt{\left((ku_1)^2 + (ku_2)^2 + (ku_3)^2\right)} = |k|\sqrt{u_1^2 + u_2^2 + u_3^2}$$

Thus the following holds.

$$\|k\vec{u}\| = |k|\,\|\vec{u}\|$$

In other words, multiplication by a scalar magnifies or shrinks the length of the vector by a factor of $|k|$. If $|k| > 1$, the length of the resulting vector will be magnified. If $|k| < 1$, the length of the resulting vector will shrink. Remember that by the definition of the absolute value, $|k| > 0$.

What about the direction? Draw a picture of $\vec{u}$ and $k\vec{u}$ where k is negative. Notice that this causes the resulting vector to point in the opposite direction while if $k > 0$ it preserves the direction the vector points. Therefore the direction can either reverse, if $k < 0$, or remain preserved, if $k > 0$.

Consider the following example.

Example 4.2.5: Graphing Scalar Multiplication

Consider the vectors $\vec{u}$ and $\vec{v}$ drawn below.



Figure 4.2.11: Diagram illustrates two vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^2$. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Draw $-\vec{u}$, $2\vec{v}$, and $-\frac{1}{2}\vec{v}$.

Solution

In order to find $-\vec{u}$, we preserve the length of $\vec{u}$ and simply reverse the direction. For $2\vec{v}$, we double the length of $\vec{v}$, while preserving the direction. Finally $-\frac{1}{2}\vec{v}$ is found by taking half the length of $\vec{v}$ and reversing the direction. These vectors are shown in the following diagram.

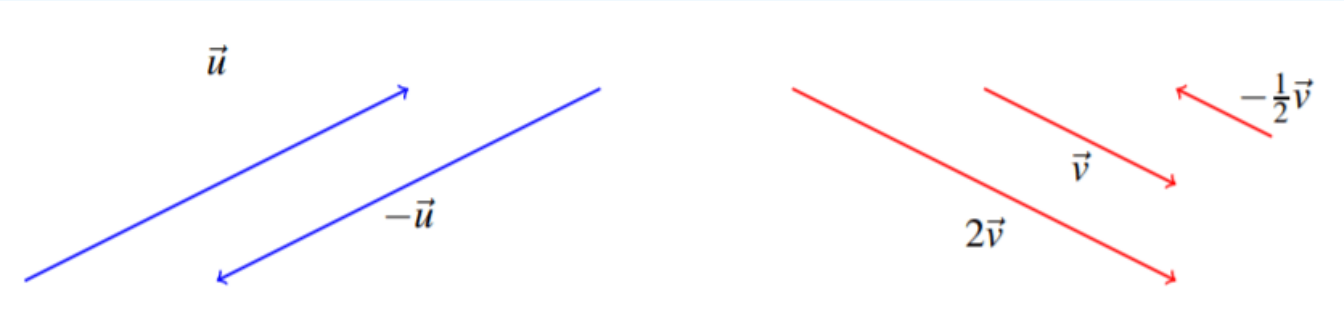


Figure 4.2.12: Diagram illustrating scalar multiplication of vectors in $\mathbb{R}^2$. The image demonstrates how to scale a vector by multiplying it by a scalar. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

In the following example, we examine the geometric meaning of this concept.

Example 4.2.6: Graphing a Linear Combination of Vectors

Consider the following picture of the vectors $\vec{u}$ and $\vec{v}$:



Figure 4.2.13: Diagram illustrates two vectors (\vec{u}\) and $\vec{v}$ in $\mathbb{R}^2$. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Sketch a picture of $\vec{u}+2\vec{v}, \vec{u}-\frac{1}{2}\vec{v}$.

Solution

The two vectors are shown below.

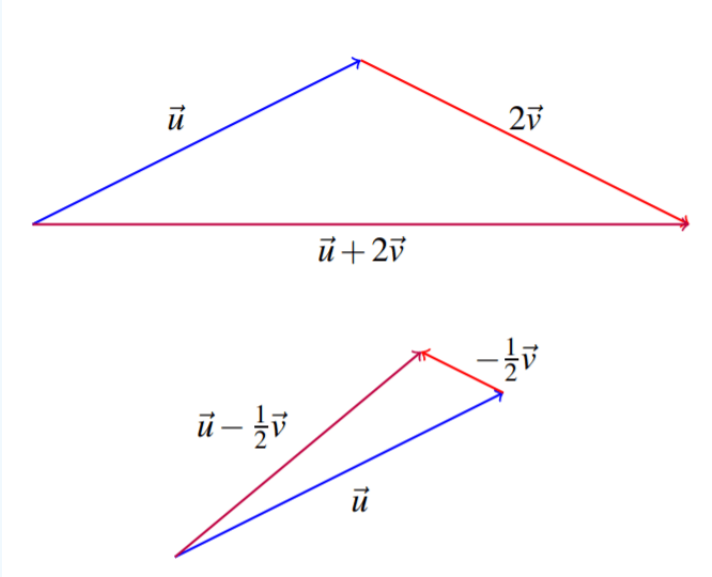


Figure 4.2.14: Diagram illustrates vector addition and scalar multiplication in $\mathbb{R}^2$. The image demonstrates how to add and scale vectors geometrically. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Scalar Multiplication of Vectors in $\mathbb{R}^n$

Scalar multiplication of vectors in $\mathbb{R}^n$ is defined as follows.

Definition 4.2.8: Scalar Multiplication of Vectors in $\mathbb{R}^n$

If $\vec{u}^{\text{int}}\mathbb{R}^n$ and $k^{\text{int}}\mathbb{R}$ is a scalar, then $k\vec{u}^{\text{int}}\mathbb{R}^n$ is defined by

$$k\vec{u} = k\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} ku_1 \\ \vdots \\ ku_n \end{bmatrix}$$

Just as with addition, scalar multiplication of vectors satisfies several important properties. These are outlined in the following theorem.

Theorem 4.2.3: Properties of Scalar Multiplication

The following properties hold for vectors $\vec{u}, \vec{v}^{\text{int}}\mathbb{R}^n$ and k, p scalars.

- The Distributive Law over Vector Addition
$$k\left(\vec{u}+\vec{v}\right)=k\vec{u}+k\vec{v}$$
- The Distributive Law over Scalar Addition
$$\left(k+p\right)\vec{u}=k\vec{u}+p\vec{u}$$
- The Associative Law for Scalar Multiplication
$$k\left(p\vec{u}\right)=\left(kp\right)\vec{u}$$
- Rule for Multiplication by 1
$$1\vec{u}=\vec{u}$$

Proof available at the end of this section.

We now present a useful notion you may have seen earlier combining vector addition and scalar multiplication

Definition 4.2.9: Linear Combination

A vector $\vec{v}$ is said to be a **linear combination** of the vectors $\vec{u}_1, \cdots, \vec{u}_n$ if there exist scalars, $a_1, \cdots, a_n$ such that
$$\vec{v}=a_1\vec{u}_1+\cdots+a_n\vec{u}_n$$

For example,
$$3\begin{bmatrix}-4\\1\\0\end{bmatrix}+2\begin{bmatrix}-3\\0\\1\end{bmatrix}=\begin{bmatrix}-18\\3\\2\end{bmatrix}.$$

Thus we can say that
$$\vec{v}=\begin{bmatrix}-18\\3\\2\end{bmatrix}$$

is a linear combination of the vectors
$$\vec{u}_1=\begin{bmatrix}-4\\1\\0\end{bmatrix} \text{ and } \vec{u}_2=\begin{bmatrix}-3\\0\\1\end{bmatrix}$$

Answers

Theorem 4.2.3: Properties of Scalar Multiplication

We will show the proof of:
$$k\left(\vec{u}+\vec{v}\right)=k\vec{u}+k\vec{v}$$

Note that:
$$\begin{aligned}k\left(\vec{u}+\vec{v}\right) &= k\left[u_1+v_1\cdots u_n+v_n\right]^T\\ &=\left[k\left(u_1+v_1\right)\cdots k\left(u_n+v_n\right)\right]^T\\ &=\left[ku_1+kv_1\cdots ku_n+kv_n\right]^T\\ &=\left[ku_1\cdots ku_n\right]^T+\left[kv_1\cdots kv_n\right]^T\\ &=k\vec{u}+k\vec{v}\end{aligned}$$

This page titled 4.2: Review of Vectors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

- 4.1: Vectors in R^n by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.
- 4.2: Vector Algebra by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.
- 4.4: Length of a Vector by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.
- 4.5: Geometric Meaning of Scalar Multiplication by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.
- 4.3: Geometric Meaning of Vector Addition by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.2E: Exercises for Section 4.1

Exercise $4.2E.1$

Find $-3\begin{bmatrix} 5 \\ -1 \\ 2 \\ -3 \end{bmatrix} + 5\begin{bmatrix} -8 \\ 2 \\ -3 \\ 6 \end{bmatrix}$.

Answer available at the end of this section.

Exercise $4.2E.2$

Find $-7\begin{bmatrix} 6 \\ 0 \\ 4 \\ -1 \end{bmatrix} + 6\begin{bmatrix} -13 \\ -1 \\ 1 \\ 6 \end{bmatrix}$.

Exercise $4.2E.3$

Let

$$\vec{a} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} -4 \\ 5 \\ 0 \end{bmatrix}.$$

Compute:

a. $\vec{a} + \vec{b}$
b. $3\vec{a}$
c. $-2\vec{b}$
d. $\vec{a} - \vec{b}$

Exercise $4.2E.4$

Let $\vec{v}$ and $\vec{w}$ be two vectors in $\mathbb{R}^2$. How is the vector sum $\vec{v} + \vec{w}$ related to the diagonal of the parallelogram formed by $\vec{v}$ and $\vec{w}$?

Exercise $4.2E.5$

Let

$$\vec{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

a. What happens to the magnitude of $\vec{v}$ when it is multiplied by 2?
b. What happens to the direction of $\vec{v}$ when it is multiplied by -1?

Answer available at the end of this section.

Exercise $4.2E.6$

Let $\vec{v}_1$ and $\vec{v}_2$ be vectors in $\mathbb{R}^2$. What geometric shape does the set $\{a\vec{v}_1 + b\vec{v}_2 \mid 0 \leq a \leq 1, 0 \leq b \leq 1\}$

Exercise $4.2E.7$

Decide whether

$$\vec{v} = \begin{bmatrix} 4 \\ 4 \\ -3 \end{bmatrix}$$

is a linear combination of the vectors

$$\vec{u}_1 = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix} \quad \text{and} \quad \vec{u}_2 = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}.$$

Answer available at the end of this section.

Exercise $4.2E.8$

Decide whether

$$\vec{v} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}$$

is a linear combination of the vectors

$$\vec{u}_1 = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix} \quad \text{and} \quad \vec{u}_2 = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}.$$

Answer available at the end of this section.

Exercise $4.2E.9$

Decide whether

$$\vec{v} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

is a linear combination of the vectors

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} 5 \\ -4 \\ 0 \end{bmatrix} \text{ and } \quad \vec{u}_3 = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}.$$

Exercise $4.2E.10$

a. Prove that there exists a unique zero vector $\vec{0}$ such that for every vector $\vec{a}^{\text{int}}\mathbb{R}^n$,

$$\vec{a} + \vec{0} = \vec{a}.$$

b. Prove that for every vector $\vec{a}^{\text{int}}\mathbb{R}^n$, there exists a unique vector $-\vec{a}$ such that \[\vec{a} + (-\vec{a}) = \vec{0}. \nonumber\

Exercise $4.2E.11$

Let $\vec{a}$ and $\vec{b}$ be two vectors in $\mathbb{R}^n$. Show geometrically that

$$\|\vec{a}+\vec{b}\|^2 + \|\vec{a}-\vec{b}\|^2 = 2\|\vec{a}\|^2 + 2\|\vec{b}\|^2.$$

Hint available at the end of this section.

Answers

Exercise $4.2E.1$

$$\begin{bmatrix} -55 \\ 13 \\ -21 \\ 39 \end{bmatrix}$$

Exercise $4.2E.5$

Exercise $4.2E.7$

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

Exercise $4.2E.8$

The system

$$\begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} = a_1\begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix} + a_2\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

has no solution.

Exercise $4.2E.11$

Sketch the parallelogram formed by the vectors $\vec{a}$ and $\vec{b}$. Notice that the two diagonal vectors of this parallelogram are given by:

$$\text{One diagonal:} \quad \vec{a}+\vec{b}$$

$$\text{Other diagonal:} \quad \vec{a}-\vec{b}$$

Then use the Law of Cosines.

This page titled 4.2E: Exercises for Section 4.1 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

4.3: Dot and Cross Product

Learning Objectives

- Compute the dot product of vectors, and use it to compute vector projections.
- Compute the cross product and the box product of vectors in $\mathbb{R}^3$.

This section explores two fundamental operations for vectors: the dot product and the cross product. These operations are widely used in applications involving vectors in $\mathbb{R}^n$.

The Dot Product

There are two ways of multiplying vectors which are of great importance in applications. The first of these is called the **dot product**. When we take the dot product of vectors, the result is a scalar. For this reason, the dot product is also called the **scalar product** and sometimes the **inner product**. The definition is as follows.

Definition 4.3.1: Dot Product

Let $\vec{u}, \vec{v}$ be two vectors in $\mathbb{R}^n$. Then we define the **dot product** $\vec{u} \bullet \vec{v}$ as

$$\vec{u} \bullet \vec{v} = \sum_{k=1}^{n} u_k v_k$$

The dot product $\vec{u} \bullet \vec{v}$ is sometimes denoted as $(\vec{u}, \vec{v})$. It can also be written as $\langle \vec{u}, \vec{v} \rangle$. If we write the vectors as column or row matrices, it is equal to the matrix product $\vec{v}\vec{w}^T$.

With this definition, there are several important properties satisfied by the dot product.

Proposition 4.3.1: Properties of the Dot Product

Let k and p denote scalars and $\vec{u}, \vec{v}, \vec{w}$ denote vectors. Then the dot product $\vec{u} \bullet \vec{v}$ satisfies the following properties.

- $\vec{u} \bullet \vec{v} = \vec{v} \bullet \vec{u}$
- $\vec{u} \bullet \vec{u} \geq 0$ and equals zero if and only if $\vec{u} = \vec{0}$
- $(k\vec{u} + p\vec{v}) \bullet \vec{w} = k(\vec{u} \bullet \vec{w}) + p(\vec{v} \bullet \vec{w})$
- $\vec{u} \bullet (k\vec{v} + p\vec{w}) = k(\vec{u} \bullet \vec{v}) + p(\vec{u} \bullet \vec{w})$
- $\|\vec{u}\|^2 = \vec{u} \bullet \vec{u}$

Proof available at the end of this section.

This proposition tells us that we can also use the dot product to find the length of a vector.

The **Cauchy Schwarz inequality** is a fundamental inequality satisfied by the dot product. It is given in the following theorem.

Theorem 4.3.1: Cauchy Schwarz Inequality

The dot product satisfies the inequality

$$\left|\vec{u} \bullet \vec{v}\right| \leq \|\vec{u}\|\|\vec{v}\| \tag{4.3.1}$$

Furthermore equality is obtained if and only if one of $\vec{u}$ or $\vec{v}$ is a scalar multiple of the other.

Proof available at the end of this section.

Notice that this proof was based only on the properties of the dot product listed in Proposition $4.3.1$. This means that whenever an operation satisfies these properties, the Cauchy Schwarz inequality holds. There are many other instances of these properties besides vectors in $\mathbb{R}^n$.

The Cauchy Schwarz inequality provides another proof of the **triangle inequality** for distances in $\mathbb{R}^n$.

Theorem $4.3.2$: Triangle Inequality

For $\vec{u}, \vec{v}^{\text{int}}\mathbb{R}^n$

$$\|\vec{u}+\vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\| \tag{4.3.2}$$

and equality holds if and only if one of the vectors is a non-negative scalar multiple of the other.

Also

$$\|\|\vec{u}\| - \|\vec{v}\|\| \leq \|\vec{u}-\vec{v}\| \tag{4.3.3}$$

Proof available at the end of this section.

The Geometric Significance of the Dot Product

Given two vectors, $\vec{u}$ and $\vec{v}$, the **included angle** is the angle between these two vectors which is given by θ such that $0 \leq \theta \leq \pi$. The dot product can be used to determine the included angle between two vectors. Consider the following picture where θ gives the included angle.

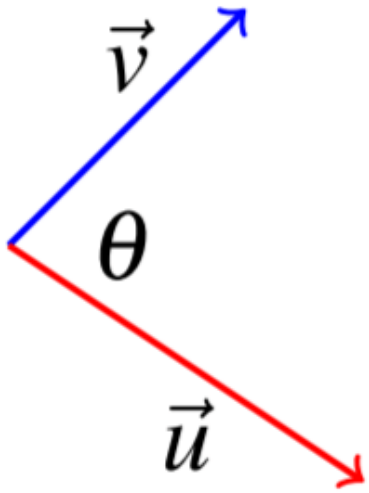


Figure $4.3.1$: Diagram illustrates the included angle θ between two vectors $\vec{u}$ and $\vec{v}$. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Proposition $4.3.2$: The Dot Product and the Included Angle

Let $\vec{u}$ and $\vec{v}$ be two vectors in $\mathbb{R}^n$, and let θ be the included angle. Then the following equation holds.

$$\vec{u} \bullet \vec{v} = \|\vec{u}\|\|\vec{v}\| \cos\theta$$

In words, the dot product of two vectors equals the product of the magnitude (or length) of the two vectors multiplied by the cosine of the included angle. Note this gives a geometric description of the dot product which does not depend explicitly on the coordinates of the vectors.

Another application of the geometric description of the dot product is in finding the angle between two lines. Typically one would assume that the lines intersect. In some situations, however, it may make sense to ask this question when the lines do not intersect, such as the angle between two object trajectories. In any case we understand it to mean the smallest angle between (any of) their direction vectors. The only subtlety here is that if $\vec{u}$ is a direction vector for a line, then so is any multiple $k\vec{u}$, and thus we will find complementary angles among all angles between direction vectors for two lines, and we simply take the smaller of the two.

Two nonzero vectors are said to be **perpendicular**, sometimes also called **orthogonal**, if the included angle is $\pi/2$ radians (90°).

Consider the following proposition.

Proposition 4.3.3: Perpendicular Vectors

Let $\vec{u}$ and $\vec{v}$ be nonzero vectors in $\mathbb{R}^n$. Then, $\vec{u}$ and $\vec{v}$ are said to be **perpendicular** exactly when

$$\vec{u} \bullet \vec{v} = 0$$

Proof available at the end of this section.

Projections

In some applications, we wish to write a vector as a sum of two related vectors. Through the concept of projections, we can find these two vectors. First, we explore an important theorem. The result of this theorem will provide our definition of a vector projection.

Theorem 4.3.3: Vector Projections

Let $\vec{v}$ and $\vec{u}$ be nonzero vectors. Then there exist unique vectors $\vec{v}_{||}$ and $\vec{v}_{\perp}$ such that

$$\vec{v} = \vec{v}_{||} + \vec{v}_{\perp} \tag{4.3.4}$$

where $\vec{v}_{||}$ is a scalar multiple of $\vec{u}$, and $\vec{v}_{\perp}$ is perpendicular to $\vec{u}$.

Proof available at the end of this section.

The vector $\vec{v}_{||}$ in Theorem 4.3.3 is called the **projection** of $\vec{v}$ onto $\vec{u}$ and is denoted by

$$\vec{v}_{||} = \text{proj}_{\vec{u}}\left(\vec{v}\right)$$

We now make a formal definition of the vector projection.

Definition 4.3.2: Vector Projection

Let $\vec{u}$ and $\vec{v}$ be vectors. Then, the **projection of $\vec{v}$ onto $\vec{u}$** is given by

$$\text{proj}_{\vec{u}}\left(\vec{v}\right) = \left(\frac{\vec{v} \bullet \vec{u}}{\vec{u} \bullet \vec{u}}\right)\vec{u} = \frac{\vec{v} \bullet \vec{u}}{\|\vec{u}\|^2}\vec{u}$$

Consider the following example of a projection.

Example 4.3.1: Find the Projection of One Vector Onto Another

Find $\text{proj}_{\vec{u}}(\vec{v})$ if

$$\vec{u} = \begin{bmatrix} 2 \\ 3 \\ -4 \end{bmatrix}, \vec{v} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Solution

We can use the formula provided in Definition 4.3.2 to find $\text{proj}_{\vec{u}}(\vec{v})$. First, compute $\vec{v} \bullet \vec{u}$. This is given by

$$\begin{aligned} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \bullet \begin{bmatrix} 2 \\ 3 \\ -4 \end{bmatrix} &= (2)(1) + (3)(-2) + (-4)(1) \\ &= 2 - 6 - 4 \\ &= -8 \end{aligned} \tag{4.3.5}$$

Similarly, $\vec{u} \bullet \vec{u}$ is given by

$$\begin{aligned} \begin{bmatrix} 2 \\ 3 \\ -4 \end{bmatrix} \bullet \begin{bmatrix} 2 \\ 3 \\ -4 \end{bmatrix} &= (2)(2) + (3)(3) + (-4)(-4) \\ &= 4 + 9 + 16 \\ &= 29 \end{aligned} \tag{4.3.6}$$

Therefore, the projection is equal to

$$\begin{aligned} \text{proj}_{\vec{u}}(\vec{v}) &= -\frac{8}{29} \begin{bmatrix} 2 \\ 3 \\ -4 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{16}{29} \\ -\frac{24}{29} \\ \frac{32}{29} \end{bmatrix} \end{aligned} \tag{4.3.7}$$

We will conclude the dot product discussion with an important application of projections. Suppose a line L and a point P are given such that P is not contained in L. Through the use of projections, we can determine the shortest distance from P to L.

Example 4.3.3: Shortest Distance from a Point to a Line

Let $P = (1, 3, 5)$ be a point in $\mathbb{R}^3$, and let L be the line which goes through point $P_0 = (0, 4, -2)$ with direction vector $\vec{d} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$. Find the shortest distance from P to the line L, and find the point Q on L that is closest to P.

Solution

In order to determine the shortest distance from P to L, we will first find the vector $\overrightarrow{P_0P}$ and then find the projection of this vector onto L. The vector $\overrightarrow{P_0P}$ is given by

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

Then, if Q is the point on L closest to P, it follows that

$$\begin{aligned}\overrightarrow{P_0Q} &= \text{proj}_{\vec{d}}\overrightarrow{P_0P} \\ &= \left(\frac{\overrightarrow{P_0P} \bullet \vec{d}}{\|\vec{d}\|^2}\right)\vec{d} \\ &= \frac{15}{9}\begin{bmatrix}2\\1\\2\end{bmatrix} \\ &= \frac{5}{3}\begin{bmatrix}2\\1\\2\end{bmatrix}\end{aligned} \tag{4.3.8}$$

Now, the distance from P to L is given by

$$\|\overrightarrow{QP}\| = \|\overrightarrow{P_0P} - \overrightarrow{P_0Q}\| = \sqrt{26}$$

The point Q is found by adding the vector $\overrightarrow{P_0Q}$ to the position vector $\overrightarrow{0P_0}$ for P_0 as follows

$$\begin{bmatrix}0\\4\\-2\end{bmatrix} + \frac{5}{3}\begin{bmatrix}2\\1\\2\end{bmatrix} = \begin{bmatrix}\frac{10}{3}\\\frac{17}{3}\\\frac{4}{3}\end{bmatrix} \tag{4.3.9}$$

Therefore, $Q = (\frac{10}{3}, \frac{17}{3}, \frac{4}{3})$.

Recall that the dot product is one of two important products for vectors. The second type of product for vectors is called the cross product. It is important to note that the cross product is only defined in $\mathbb{R}^3$. First we discuss the geometric meaning and then a description in terms of coordinates is given, both of which are important. The geometric description is essential in order to understand the applications to physics and geometry while the coordinate description is necessary to compute the cross product.

The Cross Product

Consider the following definition.

Definition 4.3.3: Right Hand System of Vectors

Three vectors, $\vec{u}, \vec{v}, \vec{w}$ form a right hand system if when you extend the fingers of your right hand along the direction of vector $\vec{u}$ and close them in the direction of $\vec{v}$, the thumb points roughly in the direction of $\vec{w}$.

For an example of a right handed system of vectors, see the following picture.

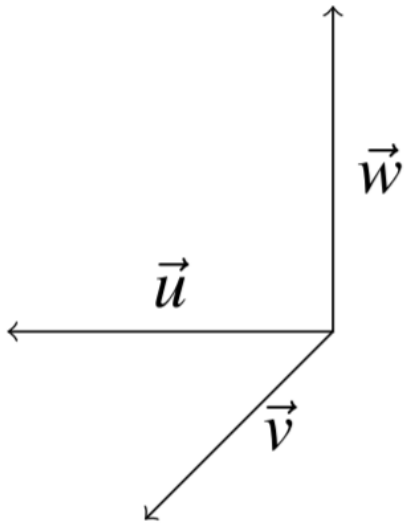


Figure 4.3.2: Diagram illustrates a right-handed coordinate system. The image shows how to determine the direction of the cross product of two vectors using the right-hand rule. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

In this picture the vector $\vec{w}$ points upwards from the plane determined by the other two vectors. Point the fingers of your right hand along $\vec{u}$, and close them in the direction of $\vec{v}$. Notice that if you extend the thumb on your right hand, it points in the direction of $\vec{w}$.

You should consider how a right hand system would differ from a left hand system. Try using your left hand and you will see that the vector $\vec{w}$ would need to point in the opposite direction.

Notice that the special vectors, $\vec{i}, \vec{j}, \vec{k}$ will always form a right handed system. If you extend the fingers of your right hand along $\vec{i}$ and close them in the direction $\vec{j}$, the thumb points in the direction of $\vec{k}$.

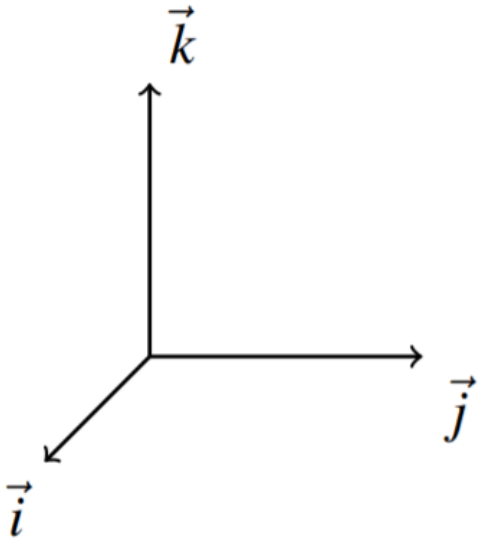


Figure 4.3.3: Diagram illustrates a coordinate system in $\mathbb{R}^3$. The image shows the standard unit vectors $\vec{i}$, $\vec{j}$, and $\vec{k}$ along the x, y, and z axes, respectively. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

The following is the geometric description of the cross product. Recall that the dot product of two vectors results in a scalar. In contrast, the cross product results in a vector, as the product gives a direction as well as magnitude.

Definition 4.3.4: Geometric Definition of Cross Product

Let $\vec{u}$ and $\vec{v}$ be two vectors in $\mathbb{R}^3$. Then the **cross product**, written $\vec{u} \times \vec{v}$, is defined by the following two rules.

1. Its length is
$$\|\vec{u} \times \vec{v}\| = \|\vec{u}\|\|\vec{v}\| \sin\theta,$$
where θ is the included angle between $\vec{u}$ and $\vec{v}$.
2. It is perpendicular to both $\vec{u}$ and $\vec{v}$, that is
$$\begin{aligned} (\vec{u} \times \vec{v}) \cdot \vec{u} &= 0, \\ (\vec{u} \times \vec{v}) \cdot \vec{v} &= 0, \end{aligned} \tag{4.3.10}$$
and
$$\vec{u}, \vec{v}, \vec{u} \times \vec{v}$$
form a right hand system.

The cross product of the special vectors $\vec{i}, \vec{j}, \vec{k}$ is as follows.

$$\begin{aligned} \vec{i} \times \vec{j} &= \vec{k} \quad \vec{j} \times \vec{i} = -\vec{k} \\ \vec{k} \times \vec{i} &= \vec{j} \quad \vec{i} \times \vec{k} = -\vec{j} \\ \vec{j} \times \vec{k} &= \vec{i} \quad \vec{k} \times \vec{j} = -\vec{i} \end{aligned}$$

With this information, the following gives the coordinate description of the cross product.

Recall that the vector $\vec{u} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}^T$ can be written in terms of $\vec{i}, \vec{j}, \vec{k}$ as $\vec{u} = u_1\vec{i} + u_2\vec{j} + u_3\vec{k}$.

Theorem 4.3.4: Coordinate Description of Cross Product

Let $\vec{u} = u_1\vec{i} + u_2\vec{j} + u_3\vec{k}$ and $\vec{v} = v_1\vec{i} + v_2\vec{j} + v_3\vec{k}$ be two vectors. Then

$$\vec{u} \times \vec{v} = (u_2v_3 - u_3v_2)\,\vec{i} - (u_1v_3 - u_3v_1)\,\vec{j} + (u_1v_2 - u_2v_1)\,\vec{k} \tag{4.3.11}$$

Writing $\vec{u} \times \vec{v}$ in the usual way, it is given by

$$\vec{u} \times \vec{v} = \begin{bmatrix} u_2v_3 - u_3v_2 \\ -(u_1v_3 - u_3v_1) \\ u_1v_2 - u_2v_1 \end{bmatrix}$$

We now prove this proposition.

Proof available at the end of this section.

There is another version of (4.3.11) which may be easier to remember. We can express the cross product as the determinant of a matrix, as follows.

$$\vec{u} \times \vec{v} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \tag{4.3.12}$$

Expanding the determinant along the top row yields

$$\vec{i}\,(-1)^{1+1}\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} + \vec{j}\,(-1)^{2+1}\begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} + \vec{k}\,(-1)^{3+1}\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}$$

\

$$= \vec{i} \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} - \vec{j} \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} + \vec{k} \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}$$

Expanding these determinants leads to

$$(u_2v_3 - u_3v_2)\,\vec{i} - (u_1v_3 - u_3v_1)\,\vec{j} + (u_1v_2 - u_2v_1)\,\vec{k}$$

which is the same as (4.3.20).

The cross product satisfies the following properties.

Proposition 4.3.4: Properties of the Cross Product

Let $\vec{u}, \vec{v}, \vec{w}$ be vectors in $\mathbb{R}^3$, and k a scalar. Then, the following properties of the cross product hold.

1. $\vec{u} \times \vec{v} = -(\vec{v} \times \vec{u})$, and $\vec{u} \times \vec{u} = \vec{0}$
2. $(k\vec{u}) \times \vec{v} = k(\vec{u} \times \vec{v}) = \vec{u} \times (k\vec{v})$
3. $\vec{u} \times (\vec{v} + \vec{w}) = \vec{u} \times \vec{v} + \vec{u} \times \vec{w}$
4. $(\vec{v} + \vec{w}) \times \vec{u} = \vec{v} \times \vec{u} + \vec{w} \times \vec{u}$

Proof available at the end of this section.

An important geometrical application of the cross product is as follows. The size of the cross product, $\|\vec{u} \times \vec{v}\|$, is the area of the parallelogram determined by $\vec{u}$ and $\vec{v}$, as shown in the following picture.

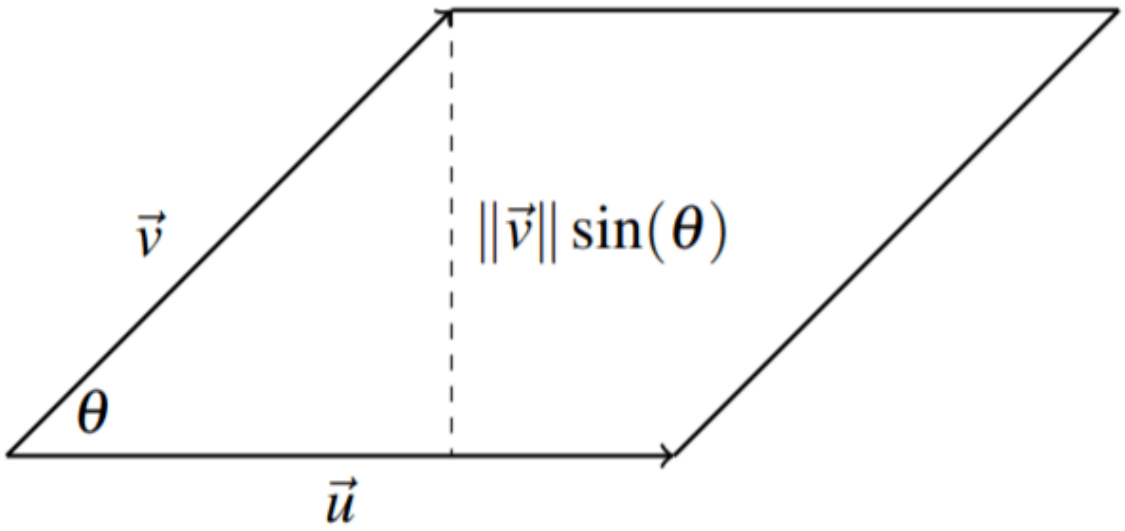


Figure 4.3.4: Diagram illustrating the geometric interpretation of the cross product of two vectors $\vec{u}$ and $\vec{v}$. The image shows the parallelogram formed by the two vectors, and the magnitude of the cross product is equal to the area of the parallelogram. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

We examine this concept in the following example.

Example 4.3.3: Area of a Parallelogram

Find the area of the parallelogram determined by the vectors $\vec{u}$ and $\vec{v}$ given by

$$\vec{u} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \vec{v} = \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix}$$

Solution

Notice that these vectors are the same as the ones given in Example 4.3.1. Recall from the geometric description of the cross product, that the area of the parallelogram is simply the magnitude of $\vec{u} \times \vec{v}$. From Example 4.3.1, $\vec{u} \times \vec{v} = 3\vec{i} + 5\vec{j} + \vec{k}$. We can also write this as

$$\vec{u} \times \vec{v} = \begin{bmatrix} 3 \\ 5 \\ 1 \end{bmatrix}$$

Thus the area of the parallelogram is

$$\|\vec{u} \times \vec{v}\| = \sqrt{(3)(3) + (5)(5) + (1)(1)} = \sqrt{9 + 25 + 1} = \sqrt{35}$$

We can also use this concept to find the area of a triangle. Consider the following example.

Example 4.3.4: Area of Triangle

Find the area of the triangle determined by the points $(1, 2, 3), (0, 2, 5), (5, 1, 2)$

Solution

This triangle is obtained by connecting the three points with lines. Picking $(1, 2, 3)$ as a starting point, there are two displacement vectors, $\begin{bmatrix} -1 & 0 & 2 \end{bmatrix}^T$ and $\begin{bmatrix} 4 & -1 & -1 \end{bmatrix}^T$. Notice that if we add either of these vectors to the position vector of the starting point, the result is the position vectors of the other two points. Now, the area of the triangle is half the area of the parallelogram determined by $\begin{bmatrix} -1 & 0 & 2 \end{bmatrix}^T$ and $\begin{bmatrix} 4 & -1 & -1 \end{bmatrix}^T$. The required cross product is given by

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

Taking the size of this vector gives the area of the parallelogram, given by

$$\sqrt{(2)(2) + (7)(7) + (1)(1)} = \sqrt{4 + 49 + 1} = \sqrt{54}$$

Hence the area of the triangle is $\frac{1}{2}\sqrt{54} = \frac{3}{2}\sqrt{6}$.

In general, if you have three points in $\mathbb{R}^3$, P, Q, R, the area of the triangle is given by

$$\frac{1}{2}\|\vec{PQ} \times \vec{PR}\|$$

Recall that $\vec{PQ}$ is the vector running from point P to point Q.

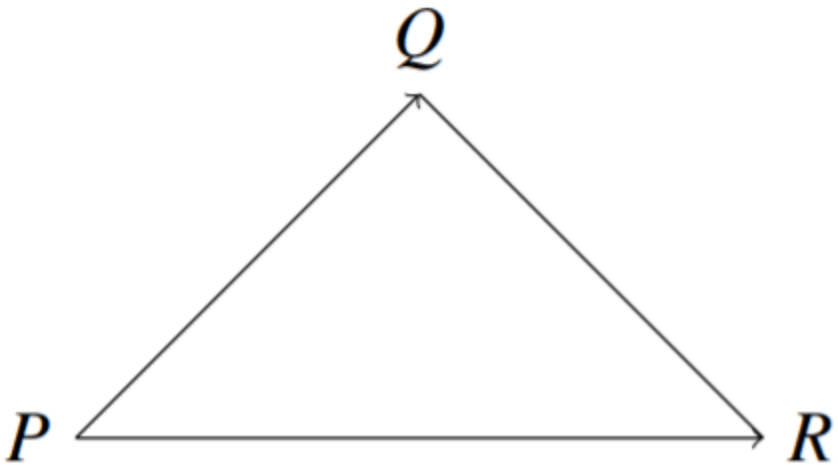


Figure 4.3.5: Diagram illustrates a triangle formed by three points P, Q, and R in $\mathbb{R}^3$. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

In the next section, we explore another application of the cross product.

The Box Product

Recall that we can use the cross product to find the the area of a parallelogram. It follows that we can use the cross product together with the dot product to find the volume of a parallelepiped. We begin with a definition.

Definition 4.3.5: Parallelepiped

A **parallelepiped** determined by the three vectors, $\vec{u}, \vec{v}$, and $\vec{w}$ consists of

$$\left\{r\vec{u} + s\vec{v} + t\vec{w} : r, s, t \in [0,1]\right\}$$

That is, if you pick three numbers, r, s, and t each in $[0,1]$ and form $r\vec{u} + s\vec{v} + t\vec{w}$ then the collection of all such points makes up the parallelepiped determined by these three vectors.

The following is an example of a parallelepiped.

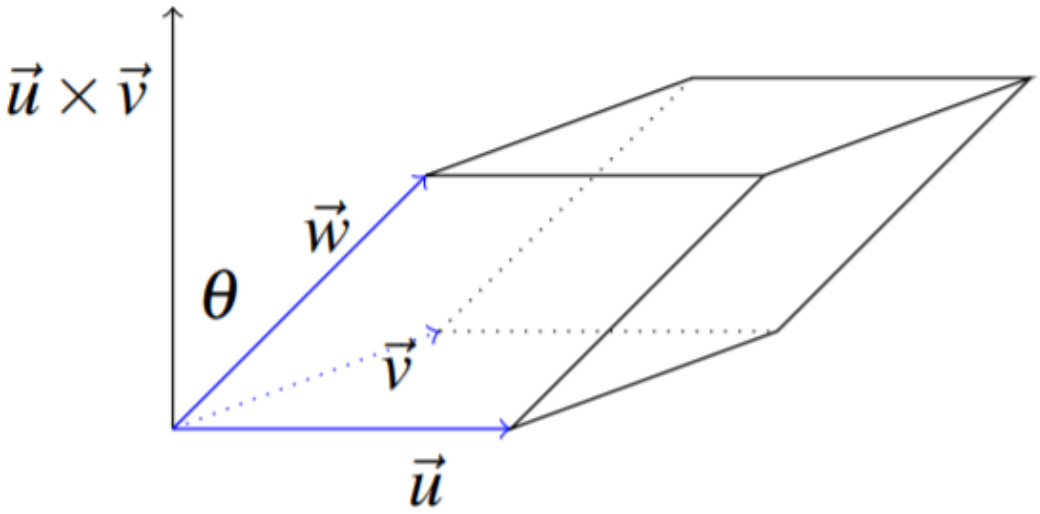


Figure 4.3.6: Diagram illustrates a parallelepiped determined by three vectors $\vec{u}$, $\vec{v}$, and $\vec{w}$. The image shows the three-dimensional shape formed by the vectors. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Notice that the base of the parallelepiped is the parallelogram determined by the vectors $\vec{u}$ and $\vec{v}$. Therefore, its area is equal to $\|\vec{u} \times \vec{v}\|$. The height of the parallelepiped is $\|\vec{w}\| \cos\theta$ where θ is the angle shown in the picture between $\vec{w}$ and $\vec{u} \times \vec{v}$. The volume of this parallelepiped is the area of the base times the height which is just

$$\|\vec{u} \times \vec{v}\| \|\vec{w}\| \cos\theta = (\vec{u} \times \vec{v}) \cdot \vec{w}$$

This expression is known as the **box product** and is sometimes written as $[\vec{u}, \vec{v}, \vec{w}]$. You should consider what happens if you interchange the $\vec{v}$ with the $\vec{w}$ or the $\vec{u}$ with the $\vec{w}$. You can see geometrically from drawing pictures that this merely introduces a minus sign. In any case the box product of three vectors always equals either the volume of the parallelepiped determined by the three vectors or else -1 times this volume.

Note: Connection to Determinants

Recall from Section 3.4 that the determinant is useful in computing the area of a parallelogram formed by two vectors and the volume of the parallelepiped formed by three vectors.

- In $\mathbb{R}^2$: Suppose $\vec{v}_1 = \langle a, b \rangle$ and $\vec{v}_2 = \langle c, d \rangle$. Then, the area of the parallelogram formed by $\vec{v}_1$ and $\vec{v}_2$ is simply the magnitude of the cross product of the $\mathbb{R}^3$ vectors $\langle a, b, 0 \rangle$ and $\langle c, d, 0 \rangle$. This is simply $|ad - bc|$, which is the absolute value of the determinant of the 2×2 matrix formed by the row vectors $\vec{v}_1$ and $\vec{v}_2$.
- In $\mathbb{R}^3$: Suppose $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_3$ are vectors in $\mathbb{R}^3$. Then, the volume of the parallelepiped formed by $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_3$ is simply the absolute value of the box product of the three vectors. We leave it to the reader to check that indeed this is also the absolute value of the determinant of the 3×3 matrix formed by the row vectors $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_3$.

Answers

Proposition 4.3.1: Properties of the Dot Product

The proof is left as an exercise.

Theorem 4.3.1: Cauchy Schwarz Inequality

First note that if $\vec{v} = \vec{0}$ both sides of (4.3.1) equal zero and so the inequality holds in this case. Therefore, it will be assumed in what follows that $\vec{v} \neq \vec{0}$.

Define a function of $t \in \mathbb{R}$ by

$$f(t) = (\vec{u} + t\vec{v}) \bullet (\vec{u} + t\vec{v})$$

Then by Proposition 4.3.1, $f(t) \geq 0$ for all $t \in \mathbb{R}$. Also from Proposition 4.3.1

$$\begin{aligned} f(t) &= \vec{u} \bullet (\vec{u} + t\vec{v}) + t\vec{v} \bullet (\vec{u} + t\vec{v}) \\ &= \vec{u} \bullet \vec{u} + t(\vec{u} \bullet \vec{v}) + t\vec{v} \bullet \vec{u} + t^2 \vec{v} \bullet \vec{v} \\ &= \|\vec{u}\|^2 + 2t(\vec{u} \bullet \vec{v}) + \|\vec{v}\|^2 t^2 \end{aligned} \tag{4.3.13}$$

Now this means the graph of $y = f(t)$ is a parabola which opens up and either its vertex touches the t axis or else the entire graph is above the t axis. In the first

case, there exists some t where $f(t)=0$ and this requires $\vec{u}+t\vec{v}=\vec{0}$ so one vector is a multiple of the other. Then clearly equality holds in (4.3.1). In the case where $\vec{v}$ is not a multiple of $\vec{u}$, it follows $f(t)>0$ for all t which says $f(t)$ has no real zeros and so from the quadratic formula,

$$\left(2\left(\vec{u}\bullet\vec{v}\right)\right)^2-4\|\vec{u}\|^2\|\vec{v}\|^2<0$$

which is equivalent to $\left|\vec{u}\bullet\vec{v}\right|<\|\vec{u}\|\|\vec{v}\|$.

Theorem 4.3.2**: Triangle Inequality**

By properties of the dot product and the Cauchy Schwarz inequality,

$$\begin{aligned}\|\vec{u}+\vec{v}\|^2&=\left(\vec{u}+\vec{v}\right)\bullet\left(\vec{u}+\vec{v}\right)\\&=\left(\vec{u}\bullet\vec{u}\right)+\left(\vec{u}\bullet\vec{v}\right)+\left(\vec{v}\bullet\vec{u}\right)+\left(\vec{v}\bullet\vec{v}\right)\\&=\|\vec{u}\|^2+2\left(\vec{u}\bullet\vec{v}\right)+\|\vec{v}\|^2\\&\leq\|\vec{u}\|^2+2\left|\vec{u}\bullet\vec{v}\right|+\|\vec{v}\|^2\\&\leq\|\vec{u}\|^2+2\|\vec{u}\|\|\vec{v}\|+\|\vec{v}\|^2=\left(\|\vec{u}\|+\|\vec{v}\|\right)^2\end{aligned}\tag{4.3.14}$$

Hence,

$$\|\vec{u}+\vec{v}\|^2\leq\left(\|\vec{u}\|+\|\vec{v}\|\right)^2$$

Taking square roots of both sides you obtain (4.3.2).

It remains to consider when equality occurs. Suppose $\vec{u}=\vec{0}$. Then, $\vec{u}=0\vec{v}$ and the claim about when equality occurs is verified. The same argument holds if $\vec{v}=\vec{0}$. Therefore, it can be assumed both vectors are nonzero. To get equality in (4.3.2) above, Theorem 4.3.1 implies one of the vectors must be a multiple of the other. Say $\vec{v}=k\vec{u}$. If $k<0$ then equality cannot occur in (4.3.2) because in this case

$$\vec{u}\bullet\vec{v}=k\|\vec{u}\|^2<0<|k|\,\|\vec{u}\|^2=\left|\vec{u}\bullet\vec{v}\right|$$

Therefore, $k\geq0$.

To get the other form of the triangle inequality write

$$\vec{u}=\vec{u}-\vec{v}+\vec{v}$$

so

$$\begin{aligned}\|\vec{u}\|&=\|\vec{u}-\vec{v}+\vec{v}\|\\&\leq\|\vec{u}-\vec{v}\|+\|\vec{v}\|\end{aligned}\tag{4.3.15}$$

Therefore,

$$\|\vec{u}\|-\|\vec{v}\|\leq\|\vec{u}-\vec{v}\|\tag{4.3.16}$$

Similarly,

$$\|\vec{v}\|-\|\vec{u}\|\leq\|\vec{v}-\vec{u}\|=\|\vec{u}-\vec{v}\|\tag{4.3.17}$$

It follows from (4.3.16) and (4.3.17) that (4.3.3) holds. This is because $\left|\|\vec{u}\|-\|\vec{v}\|\right|$ equals the left side of either (4.3.16) or (4.3.17) and either way, $\left|\|\vec{u}\|-\|\vec{v}\|\right|\leq\|\vec{u}-\vec{v}\|$.

Proposition 4.3.3**: Perpendicular Vectors**

This follows directly from Proposition 4.3.2. First if the dot product of two nonzero vectors is equal to 0, this tells us that $\cos\theta=0$ (this is where we need nonzero vectors). Thus $\theta=\pi/2$ and the vectors are perpendicular.

If on the other hand $\vec{v}$ is perpendicular to $\vec{u}$, then the included angle is $\pi/2$ radians. Hence $\cos\theta=0$ and $\vec{u}\bullet\vec{v}=0$.

Theorem 4.3.3**: Vector Projections**

Suppose (4.3.4) holds and $\vec{v}_{||}=k\vec{u}$. Taking the dot product of both sides of (4.3.4) with $\vec{u}$ and using $\vec{v}_{\perp}\bullet\vec{u}=0$, this yields

$$\begin{aligned}\vec{v}\bullet\vec{u}&=\left(\vec{v}_{||}+\vec{v}_{\perp}\right)\bullet\vec{u}\\&=k\vec{u}\bullet\vec{u}+\vec{v}_{\perp}\bullet\vec{u}\\&=k\|\vec{u}\|^2\end{aligned}$$

which requires $k=\vec{v}\bullet\vec{u}/\|\vec{u}\|^2$. Thus there can be no more than one vector $\vec{v}_{||}$. It follows $\vec{v}_{\perp}$ must equal $\vec{v}-\vec{v}_{||}$. This verifies there can be no more than one choice for both $\vec{v}_{||}$ and $\vec{v}_{\perp}$ and proves their uniqueness.

Now let

$$\vec{v}_{||}=\frac{\vec{v}\bullet\vec{u}}{\|\vec{u}\|^2}\vec{u}$$

and let

$$\vec{v}_{\perp}=\vec{v}-\vec{v}_{||}=\vec{v}-\frac{\vec{v}\bullet\vec{u}}{\|\vec{u}\|^2}\vec{u}$$

Then $\vec{v}_{||}=k\vec{u}$ where $k=\frac{\vec{v}\bullet\vec{u}}{\|\vec{u}\|^2}$. It only remains to verify $\vec{v}_{\perp}\bullet\vec{u}=0$. But

$$\begin{aligned}\vec{v}_{\perp}\bullet\vec{u}&=\vec{v}\bullet\vec{u}-\frac{\vec{v}\bullet\vec{u}}{\|\vec{u}\|^2}\vec{u}\bullet\vec{u}\\&=\vec{v}\bullet\vec{u}-\vec{v}\bullet\vec{u}\\&=0\end{aligned}\tag{4.3.18}$$

Theorem 4.3.4**: Coordinate Description of Cross Product**

From the above table and the properties of the cross product listed,

$$\begin{aligned}\vec{u}\times\vec{v} &= \left(u_1\vec{i}+u_2\vec{j}+u_3\vec{k}\right)\times\left(v_1\vec{i}+v_2\vec{j}+v_3\vec{k}\right)\\ &= u_1v_2\vec{i}\times\vec{j}+u_1v_3\vec{i}\times\vec{k}+u_2v_1\vec{j}\times\vec{i}+u_2v_3\vec{j}\times\vec{k}++u_3v_1\vec{k}\times\vec{i}+u_3v_2\vec{k}\times\vec{j} \qquad (4.3.19)\\ &= u_1v_2\vec{k}-u_1v_3\vec{j}-u_2v_1\vec{k}+u_2v_3\vec{i}+u_3v_1\vec{j}-u_3v_2\vec{i}\\ &= (u_2v_3-u_3v_2)\vec{i}+(u_3v_1-u_1v_3)\vec{j}+(u_1v_2-u_2v_1)\vec{k} \qquad (4.3.20)\end{aligned}$$

Proposition $4.3.4$**: Properties of the Cross Product**

Formula $1.$ follows immediately from the definition. The vectors $\vec{u}\times\vec{v}$ and $\vec{v}\times\vec{u}$ have the same magnitude, $|\vec{u}|\,|\vec{v}|\sin\theta$, and an application of the right hand rule shows they have opposite direction.

Formula $2.$ is proven as follows. If k is a non-negative scalar, the direction of $(k\vec{u})\times\vec{v}$ is the same as the direction of $\vec{u}\times\vec{v}, k(\vec{u}\times\vec{v})$ and $\vec{u}\times(k\vec{v})$. The magnitude is k times the magnitude of $\vec{u}\times\vec{v}$ which is the same as the magnitude of $k(\vec{u}\times\vec{v})$ and $\vec{u}\times(k\vec{v})$. Using this yields equality in $2.$ In the case where $k<0$, everything works the same way except the vectors are all pointing in the opposite direction and you must multiply by $|k|$ when comparing their magnitudes.

The distributive laws, $3.$ and $4.$, are much harder to establish. For now, it suffices to notice that if we know that $3.$ is true, $4.$ follows. Thus, assuming $3.$, and using $1.$,

$$\begin{aligned}(\vec{v}+\vec{w})\times\vec{u} &= -\vec{u}\times(\vec{v}+\vec{w})\\ &= -(\vec{u}\times\vec{v}+\vec{u}\times\vec{w}) \qquad (4.3.21)\\ &= \vec{v}\times\vec{u}+\vec{w}\times\vec{u}\end{aligned}$$

This page titled 4.3: Dot and Cross Product is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

- 4.7: The Dot Product by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.
- 4.9: The Cross Product by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.3E: Exercises for Section 4.2

Exercise $4.3E.1$

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

Answer available at the end of this section.

Exercise $4.3E.2$

Use the formula given in Proposition 4.2.1 to verify the Cauchy Schwarz inequality and to show that equality occurs if and only if one of the vectors is a scalar multiple of the other

Answer available at the end of this section.

Exercise $4.3E.3$

Let $\vec{a}$, $\vec{b}$ be vectors. Show that $\left(\vec{a} \bullet \vec{b}\right) = \frac{1}{4}\left(||\vec{a}+\vec{b}||^2 - ||\vec{a}-\vec{b}||^2\right)$.

Exercise $4.3E.4$

Using the axioms of the dot product, prove the parallelogram identity:

$$||\vec{a}+\vec{b}||^2 + ||\vec{a}-\vec{b}||^2 = 2||\vec{a}||^2 + 2||\vec{b}||^2$$

Exercise $4.3E.5$

Let A be a real $m \times n$ matrix and let $\vec{u} \in \mathbb{R}^n$ and $\vec{v} \in \mathbb{R}^m$. Show $A\vec{u} \bullet \vec{v} = \vec{u} \bullet A^T\vec{v}$. Hint: Use the definition of matrix multiplication to do this.

Answer available at the end of this section.

Exercise $4.3E.6$

Use the result of Problem $4.3E.5$ to verify directly that $(AB)^T = B^TA^T$ without making any reference to subscripts.

Answer available at the end of this section.

Exercise $4.3E.7$

Find the angle between the vectors

$$\vec{u} = \begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix}, \vec{v} = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}$$

Answer available at the end of this section.

Exercise $4.3E.8$

Find the angle between the vectors

$$\vec{u} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \vec{v} = \begin{bmatrix} 1 \\ 2 \\ -7 \end{bmatrix}$$

Answer available at the end of this section.

Exercise $4.3E.9$

Find $\text{proj}_{\vec{v}}(\vec{w})$ where $\vec{w} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Answer available at the end of this section.

Exercise $4.3E.10$

Find $\text{proj}_{\vec{v}}(\vec{w})$ where $\vec{w} = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$.

Answer available at the end of this section.

Exercise $4.3E.11$

Find $\text{proj}_{\vec{v}}(\vec{w})$ where $\vec{w} = \begin{bmatrix} 1 \\ 2 \\ -2 \\ 1 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix}$.

Answer available at the end of this section.

Exercise $4.3E.12$

Let $P = (1, 2, 3)$ be a point in $\mathbb{R}^3$. Let L be the line through the point $P_0 = (1, 4, 5)$ with direction vector $\vec{d} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$. Find the shortest distance from P to L, and find the point Q on L that is closest to P.

Exercise $4.3E.13$

Let $P = (0, 2, 1)$ be a point in $\mathbb{R}^3$. Let L be the line through the point $P_0 = (1, 1, 1)$ with direction vector $\vec{d} = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$. Find the shortest distance from P to L, and find the point Q on L that is closest to P.

Exercise $4.3E.14$

Does it make sense to speak of $\text{proj}_{\vec{0}}(\vec{w})$?

Answer available at the end of this section.

Exercise $4.3E.15$

The Cauchy Schwarz inequality in $\mathbb{R}^n$ is:

For vectors $\vec{u}$, $\vec{v}$, $(\vec{u} - \text{proj}_{\vec{v}}\vec{u}) \bullet (\vec{u} - \text{proj}_{\vec{v}}\vec{u}) \geq 0$.

a. Prove the Cauchy Schwarz Inequality. HINT: Begin by simplifying using the axioms of the dot product and then putting in the formula for the projection. Notice that this expression equals 0 and you get equality in the Cauchy Schwarz inequality if and only if $\vec{u} = \text{proj}_{\vec{v}}\vec{u}$.
b. What is the geometric meaning of $\vec{w} = \text{proj}_{\vec{v}}\vec{w}$?

Answer available at the end of this section.

Exercise $4.3E.16$

Let $\vec{v}$, $\vec{w}$, $\vec{u}$ be vectors. Show that $(\vec{w} + \vec{u})_\perp = \vec{w}_\perp + \vec{u}_\perp$ where $\vec{w}_\perp = \vec{w} - \text{proj}_{\vec{v}}(\vec{w})$.

Answer available at the end of this section.

Exercise $4.3E.17$

Show that

$$(\vec{v} - \text{proj}_{\vec{u}}(\vec{v}), \vec{u}) = (\vec{v} - \text{proj}_{\vec{u}}(\vec{v})) \bullet \vec{u} = 0$$

and conclude every vector in $\mathbb{R}^n$ can be written as the sum of two vectors, one which is perpendicular and one which is parallel to the given vector.

Answer available at the end of this section.

Exercise $4.3E.18$

Show that if $\vec{a} \times \vec{u} = \vec{0}$ for any unit vector $\vec{u}$, then $\vec{a} = \vec{0}$.

Answer available at the end of this section.

Exercise $4.3E.19$

Find the area of the triangle determined by the three points $(1,2,3),\ (4,2,0)$ and $(-3,2,1)$.

Answer available at the end of this section.

Exercise $4.3E.20$

Find the area of the triangle determined by the three points $(1,0,3),\ (4,1,0)$ and $(-3,1,1)$.

Answer available at the end of this section.

Exercise $4.3E.21$

Find the area of the triangle determined by the three points, $(1,2,3),\ (2,3,4)$ and $(3,4,5)$. Did something interesting happen here? What does it mean geometrically?

Answer available at the end of this section.

Exercise $4.3E.22$

Find the area of the parallelogram determined by the vectors $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix}$.

Answer available at the end of this section.

Exercise $4.3E.23$

Find the area of the parallelogram determined by the vectors $\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix}$.

Answer available at the end of this section.

Exercise $4.3E.24$

Is $\vec{u} \times (\vec{v} \times \vec{w}) = (\vec{u} \times \vec{v}) \times \vec{w}$? What is the meaning of $\vec{u} \times \vec{v} \times \vec{w}$? Explain. Hint: Try $\left(\vec{i} \times \vec{j}\right) \times \vec{k}$.

Answer available at the end of this section.

Exercise $4.3E.25$

Verify directly that the coordinate description of the cross product, $\vec{u} \times \vec{v}$ has the property that it is perpendicular to both $\vec{u}$ and $\vec{v}$. Then show by direct computation that this coordinate description satisfies

$$\begin{aligned} ||\vec{u} \times \vec{v}||^2 &= ||\vec{u}||^2||\vec{v}||^2 - (\vec{u} \bullet \vec{v})^2 \\ &= ||\vec{u}||^2||\vec{v}||^2(1 - \cos^2(\theta)) \end{aligned} \tag{4.3E.1}$$

where θ is the angle included between the two vectors. Explain why $||\vec{u} \times \vec{v}||$ has the correct magnitude.

Answer available at the end of this section.

Exercise $4.3E.26$

Suppose A is a 3×3 skew symmetric matrix such that $A^T = -A$. Show there exists a vector $\vec{\Omega}$ such that for all $\vec{u} \in \mathbb{R}^3$

$$A\vec{u} = \vec{\Omega} \times \vec{u}$$

Hint: Explain why since A is skew symmetric it is of the form

$$A = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}$$

where the ω_i are numbers. Then consider $\omega_1\vec{i} + \omega_2\vec{j} + \omega_3\vec{k}$.

Exercise $4.3E.27$

Find the volume of the parallelepiped determined by the vectors $\begin{bmatrix} 1 \\ -7 \\ -5 \end{bmatrix}$, $\begin{bmatrix} 1 \\ -2 \\ -6 \end{bmatrix}$, and $\begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}$.

Answer available at the end of this section.

Exercise $4.3E.28$

Suppose $\vec{u}$, $\vec{v}$, and $\vec{w}$ are three vectors whose components are all integers. Can you conclude the volume of the parallelepiped determined from these three vectors will always be an integer?

Answer available at the end of this section.

Exercise $4.3E.29$

What does it mean geometrically if the box product of three vectors gives zero?

Answer available at the end of this section.

Answers

Exercise $4.3E.1$

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \bullet \begin{bmatrix} 2 \\ 0 \\ 1 \\ 3 \end{bmatrix} = 17$$

Exercise $4.3E.2$

This formula says that $\vec{u} \bullet \vec{v} = ||\vec{u}||\ ||\vec{v}|| \cos\theta$ where θ is the included angle between the two vectors. Thus

$$||\vec{u} \bullet \vec{v}|| = ||\vec{u}||\ ||\vec{v}||\ |\cos\theta|| \leq ||\vec{u}||\ ||\vec{v}||$$

and equality holds if and only if $\theta = 0$ or π. This means that the two vectors either point in the same direction or opposite directions. Hence one is a multiple of the other.

Exercise $4.3E.5$

$A\vec{x} \bullet \vec{y} = \sum_k (A\vec{x})_k y_k = \sum_k \sum_i A_{ki} x_i y_k = \sum_i \sum_k A^T_{ik} x_i y_k = \vec{x} \bullet A^T\vec{y}$

Exercise $4.3E.6$

$$\begin{aligned} AB\vec{x} \bullet \vec{y} &= B\vec{x} \bullet A^T\vec{y} \\ &= \vec{x} \bullet B^T A^T \vec{y} \\ &= \vec{x} \bullet (AB)^T \vec{y} \end{aligned} \tag{4.3E.2}$$

Since this is true for all $\vec{x}$, it follows that, in particular, it holds for

$$\vec{x} = B^T A^T \vec{y} - (AB)^T \vec{y}$$

and so from the axioms of the dot product,

$$\left(B^T A^T \vec{y} - (AB)^T \vec{y}\right) \bullet \left(B^T A^T \vec{y} - (AB)^T \vec{y}\right) = 0$$

and so $B^T A^T \vec{y} - (AB)^T \vec{y} = \vec{0}$. However, this is true for all $\vec{y}$ and so $B^T A^T - (AB)^T = 0$.

Exercise $4.3E.7$

$\frac{\begin{bmatrix} 3 & -1 & -1 \end{bmatrix}^T \bullet \begin{bmatrix} 1 & 4 & 2 \end{bmatrix}^T}{\sqrt{9+1+1}\sqrt{1+16+4}} = \frac{-3}{\sqrt{11}\sqrt{21}} = -0.19739 = \cos\theta$ Therefore we need to solve

$$-0.19739 = \cos\theta$$

Thus $\theta = 1.7695$ radians.

Exercise $4.3E.8$

$\frac{-10}{\sqrt{1+4+1}\sqrt{1+4+49}} = -0.55555 = \cos\theta$ Therefore we need to solve $-0.55555 = \cos\theta$, which gives $\theta = 2.0313$ radians.

Exercise $4.3E.9$

$\frac{\vec{u}\bullet\vec{v}}{\vec{u}\bullet\vec{u}}\vec{u} = \frac{-5}{14}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -\frac{5}{14} \\ -\frac{5}{7} \\ -\frac{15}{14} \end{bmatrix}$

Exercise $4.3E.10$

$\frac{\vec{u}\bullet\vec{v}}{\vec{u}\bullet\vec{u}}\vec{u} = \frac{-5}{10}\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ -\frac{3}{2} \end{bmatrix}$

Exercise $4.3E.11$

$\frac{\vec{u}\bullet\vec{v}}{\vec{u}\bullet\vec{u}}\vec{u} = \frac{\begin{bmatrix} 1 & 2 & -2 & 1 \end{bmatrix}^T \bullet \begin{bmatrix} 1 & 2 & 3 & 0 \end{bmatrix}^T}{1+4+9}\begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{14} \\ -\frac{1}{7} \\ -\frac{3}{14} \\ 0 \end{bmatrix}$

Exercise $4.3E.14$

No, it does not. The 0 vector has no direction. The formula for $\text{proj}_{\vec{0}}(\vec{w})$ doesn't make sense either.

Exercise $4.3E.15$

a. $$\left(\vec{u} - \frac{\vec{u}\bullet\vec{v}}{||\vec{v}||^2}\vec{v}\right) \bullet \left(\vec{u} - \frac{\vec{u}\bullet\vec{v}}{||\vec{v}||^2}\vec{v}\right) = ||\vec{u}||^2 - 2(\vec{u}\bullet\vec{v})^2\frac{1}{||\vec{v}||^2} + (\vec{u}\bullet\vec{v})^2\frac{1}{||\vec{v}||^2} \geq 0$$

And so

$$||\vec{u}||^2||\vec{v}||^2 \geq (\vec{u}\bullet\vec{v})^2$$

You get equality exactly when $\vec{u} = \text{proj}_{\vec{v}}\vec{u} = \frac{\vec{u}\bullet\vec{v}}{||\vec{v}||^2}\vec{v}$.

b. Equality indicates that $\vec{u}$ is a multiple of $\vec{v}$.

Exercise $4.3E.16$

$$\begin{aligned} \vec{w} - \text{proj}_{\vec{v}}(\vec{w}) + \vec{u} - \text{proj}_{\vec{v}}(\vec{u}) &= \vec{w} + \vec{u} - (\text{proj}_{\vec{v}}(\vec{w}) + \text{proj}_{\vec{v}}(\vec{u})) \\ &= \vec{w} + \vec{u} - \text{proj}_{\vec{v}}(\vec{w} + \vec{u}) \end{aligned} \tag{4.3E.3}$$

This follows because

$$\begin{aligned} \text{proj}_{\vec{v}}(\vec{w}) + \text{proj}_{\vec{v}}(\vec{u}) &= \frac{\vec{u}\bullet\vec{v}}{||\vec{v}||^2}\vec{v} + \frac{\vec{w}\bullet\vec{v}}{||\vec{v}||^2}\vec{v} \\ &= \frac{(\vec{u}+\vec{w})\bullet\vec{v}}{||\vec{v}||^2}\vec{v} \\ &= \text{proj}_{\vec{v}}(\vec{w}+\vec{u}) \end{aligned} \tag{4.3E.4}$$

Exercise $4.3E.17$

$(\vec{v} - \text{proj}_{\vec{u}}(\vec{v})) \bullet \vec{u} = \vec{v}\bullet\vec{u} - \left(\frac{(\vec{v}\cdot\vec{u}}{||\vec{u}||^2}\vec{u}\right) \bullet \vec{u} = \vec{v}\bullet\vec{u} - \vec{v}\bullet\vec{u} = 0$. Therefore, $\vec{v} = \vec{v} - \text{proj}_{\vec{u}}(\vec{v}) + \text{proj}_{\vec{u}}(\vec{v})$. The first is perpendicular to $\vec{u}$ and the second is a multiple of $\vec{u}$ so it is parallel to $\vec{u}$.

Exercise $4.3E.18$

If $\vec{a} \neq \vec{0}$, then the condition says that $||\vec{a}\times\vec{u}|| = ||\vec{a}||\sin\theta = 0$ for all angles θ. Hence $\vec{a} = \vec{0}$ after all.

Exercise $4.3E.19$

$\begin{bmatrix} 3 \\ 0 \\ -3 \end{bmatrix} \times \begin{bmatrix} -4 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 18 \\ 0 \end{bmatrix}$. So the area is 9.

Exercise $4.3E.20$

$\begin{bmatrix} 3 \\ 1 \\ -3 \end{bmatrix} \times \begin{bmatrix} -4 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ 18 \\ 7 \end{bmatrix}$. The area is given by

$$\frac{1}{2}\sqrt{1+(18)^2+49} = \frac{1}{2}\sqrt{374}$$

Exercise $4.3E.21$

$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 2 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$. The area is 0. It means the three points are on the same line.

Exercise $4.3E.22$

$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \times \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ -8 \end{bmatrix}$. The area is $8\sqrt{3}$.

Exercise $4.3E.23$

$\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} \times \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 11 \\ -2 \end{bmatrix}$. The area is $\sqrt{36+121+4} = \sqrt{161}$.

Exercise $4.3E.24$

$\left(\vec{i} \times \vec{j}\right) \times \vec{j} = \vec{k} \times \vec{j} = i\vec{i}$. However, $\vec{i} \times \left(\vec{j} \times \vec{j}\right) = \vec{0}$ and so the cross product is not associative.

Exercise $4.3E.25$

Verify directly from the coordinate description of the cross product that the right hand rule applies to the vectors $\vec{i}, \vec{j}, \vec{k}$. Next verify that the distributive law holds for the coordinate description of the cross product. This gives another way to approach the cross product. First define it in terms of coordinates and then get the geometric properties from this. However, this approach does not yield the right hand rule property very easily. From the coordinate description,

$$\vec{a} \times \vec{b} \cdot \vec{a} = \epsilon_{ijk} a_j b_k a_i = -\epsilon_{jik} a_k b_k a_i = -\epsilon_{jik} b_k a_i a_j = -\vec{a} \times \vec{b} \cdot \vec{a}$$

and so $\vec{a} \times \vec{b}$ is perpendicular to $\vec{a}$. Similarly, $\vec{a} \times \vec{b}$ is perpendicular to $\vec{b}$. Now we need that

$$||\vec{a} \times \vec{b}||^2 = ||\vec{a}||^2 ||\vec{b}||^2 (1 - \cos^2\theta) = ||\vec{a}||^2 ||\vec{b}||^2 \sin^2\theta$$

and so $||\vec{a} \times \vec{b}|| = ||\vec{a}||\ ||\vec{b}|| \sin\theta$, the area of the parallelogram determined by $\vec{a}$, $\vec{b}$. Only the right hand rule is a little problematic. However, you can see right away from the component definition that the right hand rule holds for each of the standard unit vectors. Thus $\vec{i} \times \vec{j} = \vec{k}$ etc.

$$\begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = \vec{k}$$

Exercise $4.3E.27$

$$\begin{vmatrix} 1 & -7 & -5 \\ 1 & -2 & -6 \\ 3 & 2 & 3 \end{vmatrix} = 113$$

Exercise $4.3E.28$

Yes. It will involve the sum of product of integers and so it will be an integer.

Exercise $4.3E.29$

It means that if you place them so that they all have their tails at the same point, the three will lie in the same plane.

This page titled 4.3E: Exercises for Section 4.2 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

4.4: Lines and Planes

Learning Objectives

- Find the vector and parametric equations of a line.
- Find the vector and scalar equations of a plane.

We can use the concept of vectors and points to find equations for arbitrary lines in $\mathbb{R}^n$, although in this section the focus will be on lines in $\mathbb{R}^3$.

To begin, consider the case $n = 1$ so we have $\mathbb{R}^1 = \mathbb{R}$. There is only one line here which is the familiar number line, that is $\mathbb{R}$ itself. Therefore it is not necessary to explore the case of $n = 1$ further.

Now consider the case where $n = 2$, in other words $\mathbb{R}^2$. Let P and P_0 be two different points in $\mathbb{R}^2$ which are contained in a line L. Let $\vec{p}$ and $\vec{p_0}$ be the position vectors for the points P and P_0 respectively. Suppose that Q is an arbitrary point on L. Consider the following diagram.

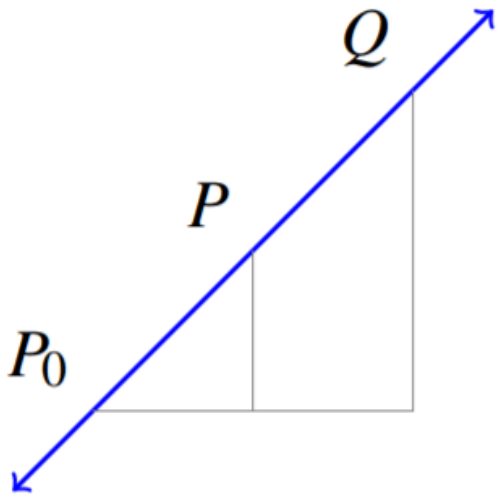


Figure 4.4.1 This diagram illustrates the concept of a line in two dimensions. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Our goal is to be able to define Q in terms of P and P_0. Consider the vector $\overrightarrow{P_0P} = \vec{p} - \vec{p_0}$ which has its tail at P_0 and point at P. If we add $\vec{p} - \vec{p_0}$ to the position vector $\vec{p_0}$ for P_0, the sum would be a vector with its point at P. In other words,

$$\vec{p} = \vec{p_0} + (\vec{p} - \vec{p_0})$$

Now suppose we were to add $t(\vec{p} - \vec{p_0})$ to $\vec{p}$ where t is some scalar. You can see that by doing so, we could find a vector with its point at Q. In other words, we can find t such that

$$\vec{q} = \vec{p_0} + t\,(\vec{p} - \vec{p_0})$$

This equation determines the line L in $\mathbb{R}^2$. In fact, it determines a line L in $\mathbb{R}^n$. Consider the following definition.

Definition 4.4.1: Vector Equation of a Line

Suppose a line L in $\mathbb{R}^n$ contains the two different points P and P_0. Let $\vec{p}$ and $\vec{p_0}$ be the position vectors of these two points, respectively. Then, L is the collection of points Q which have the position vector $\vec{q}$ given by

$$\vec{q} = \vec{p_0} + t\left(\vec{p} - \vec{p_0}\right)$$

where $t^{\text{int}}\mathbb{R}$.

Let $\vec{d} = \vec{p} - \vec{p_0}$. Then $\vec{d}$ is the **direction vector for** L and the **vector equation** for L is given by

$$\vec{p} = \vec{p_0} + t\vec{d}, t^{\text{int}}\mathbb{R}$$

Note that this definition agrees with the usual notion of a line in two dimensions and so this is consistent with earlier concepts. Consider now points in $\mathbb{R}^3$. If a point $P^{\text{int}}\mathbb{R}^3$ is given by $P = (x, y, z)$, $P_0^{\text{int}}\mathbb{R}^3$ by $P_0 = (x_0, y_0, z_0)$, then we can write

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + t \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

where $\vec{d} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. This is the vector equation of L written in **component form** .

The following theorem claims that such an equation is in fact a line.

Proposition 4.4.1: Algebraic Description of a Straight Line

Let $\vec{a}, \vec{b}^{\text{int}}\mathbb{R}^n$ with $\vec{b} \neq \vec{0}$. Then $\vec{x} = \vec{a} + t\vec{b}, \quad t^{\text{int}}\mathbb{R}$, is a line.

Proof available at the end of this section.

We can use the above discussion to find the equation of a line when given two distinct points. Consider the following example.

Example 4.4.1: A Line From Two Points

Find a vector equation for the line through the points $P_0 = (1, 2, 0)$ and $P = (2, -4, 6)$.

Solution

We will use the definition of a line given above in Definition 4.4.1 to write this line in the form

$$\vec{q} = \vec{p_0} + t\left(\vec{p} - \vec{p_0}\right)$$

Let $\vec{q} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$. Then, we can find $\vec{p}$ and $\vec{p_0}$ by taking the position vectors of points P and P_0 respectively. Then,

$$\vec{q} = \vec{p_0} + t\left(\vec{p} - \vec{p_0}\right)$$

can be written as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ -6 \\ 6 \end{bmatrix}, \ t^{\text{int}}\mathbb{R}$$

Here, the direction vector $\begin{bmatrix} 1 \\ -6 \\ 6 \end{bmatrix}$ is obtained by $\vec{p} - \vec{p_0} = \begin{bmatrix} 2 \\ -4 \\ 6 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$ as indicated above in Definition 4.4.1.

Notice that in the above example we said that we found "a" vector equation for the line, not "the" equation. The reason for this terminology is that there are infinitely many different vector equations for the same line. To see this, replace t with another parameter, say $3s$. Then you obtain a different vector equation for the same line because the same set of points is obtained.

In Example 4.4.1, the vector given by $\begin{bmatrix} 1 \\ -6 \\ 6 \end{bmatrix}$ is the direction vector defined in Definition 4.4.1. If we know the direction vector of a line, as well as a point on the line, we can find the vector equation.

Consider the following example.

Example 4.4.2: A Line From a Point and a Direction Vector

Find a vector equation for the line which contains the point $P_0 = (1, 2, 0)$ and has direction vector $\vec{d} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

Solution

We will use Definition 4.4.1 to write this line in the form $\vec{p} = \vec{p_0} + t\vec{d}, \ t^{\text{int}}\mathbb{R}$. We are given the direction vector $\vec{d}$. In order to find $\vec{p_0}$, we can use the position vector of the point P_0. This is given by $\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$. Letting $\vec{p} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, the equation for the line is given by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \ t^{\text{int}}\mathbb{R} \tag{4.4.1}$$

We sometimes elect to write a line such as the one given in (4.4.1) in the form

$$\left.\begin{array}{l} x = 1 + t \\ y = 2 + 2t \\ z = t \end{array}\right\} \text{ where } t^{\text{int}}\mathbb{R} \tag{4.4.2}$$

This set of equations give the same information as (4.4.1), and is called the **parametric equation of the line**.

Consider the following definition.

Definition 4.4.2: Parametric Equation of a Line

Let L be a line in $\mathbb{R}^3$ which has direction vector $\vec{d} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ and goes through the point $P_0 = (x_0, y_0, z_0)$. Then, letting t be a parameter, we can write L as

$$\left.\begin{array}{l} x = x_0 + ta \\ y = y_0 + tb \\ z = z_0 + tc \end{array}\right\} \text{ where } t^{\text{int}}\mathbb{R}$$

This is called a **parametric equation** of the line L.

You can verify that the form discussed following Example 4.4.2 in equation (4.4.2) is of the form given in Definition 4.4.2.

There is one other form for a line which is useful, which is the **symmetric form**. Consider the line given by (4.4.2). You can solve for the parameter t to write

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

Therefore,

$$x-1 = \frac{y-2}{2} = z$$

This is the **symmetric form** of the line.

In the following example, we look at how to take the equation of a line from symmetric form to parametric form.

Example 4.4.3: Change Symmetric Form to Parametric Form

Suppose the **symmetric form of a line** is

$$\frac{x-2}{3} = \frac{y-1}{2} = z+3$$

Write the line in parametric form as well as vector form.

Solution

We want to write this line in the form given by Definition 4.4.2. This is of the form

$$\left.\begin{aligned} x &= x_0 + ta \\ y &= y_0 + tb \\ z &= z_0 + tc \end{aligned}\right\} \text{ where } t \in \mathbb{R}$$

Let $t = \frac{x-2}{3}, t = \frac{y-1}{2}$ and $t = z+3$, as given in the symmetric form of the line. Then solving for x, y, z, yields

$$\left.\begin{aligned} x &= 2+3t \\ y &= 1+2t \\ z &= -3+t \end{aligned}\right\} \text{ with } t \in \mathbb{R}$$

This is the parametric equation for this line.

Now, we want to write this line in the form given by Definition 4.4.1. This is the form

$$\vec{p} = \vec{p_0} + t\vec{d}$$

where $t \in \mathbb{R}$. This equation becomes

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} + t \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \ t \in \mathbb{R}$$

Much like the above discussion with lines, vectors can be used to determine planes in $\mathbb{R}^n$. Given a vector $\vec{n}$ in $\mathbb{R}^n$ and a point P_0, it is possible to find a **unique** plane which contains P_0 and is perpendicular to the given vector.

Definition 4.4.3: Normal Vector

Let $\vec{n}$ be a nonzero vector in $\mathbb{R}^n$. Then $\vec{n}$ is called a **normal vector** to a plane if and only if

$$\vec{n} \bullet \vec{v} = 0$$

for every vector $\vec{v}$ in the plane.

In other words, we say that $\vec{n}$ is orthogonal (perpendicular) to every vector in the plane.

Consider now a plane with normal vector given by $\vec{n}$, and containing a point P_0. Notice that this plane is unique. If P is an arbitrary point on this plane, then by definition the normal vector is orthogonal to the vector between P_0 and P. Letting $\overrightarrow{0P}$ and $\overrightarrow{0P_0}$ be the position vectors of points P and P_0 respectively, it follows that

$$\vec{n} \bullet (\overrightarrow{0P} - \overrightarrow{0P_0}) = 0$$

or

$$\vec{n} \bullet \overrightarrow{P_0P} = 0$$

The first of these equations gives the vector equation of the plane.

Definition 4.4.4: Vector Equation of a Plane

Let $\vec{n}$ be the normal vector for a plane which contains a point P_0. If P is an arbitrary point on this plane, then the vector equation of the plane is given by

$$\vec{n} \bullet (\overrightarrow{0P} - \overrightarrow{0P_0}) = 0$$

Notice that this equation can be used to determine if a point P is contained in a certain plane.

Example 4.4.4: A Point in a Plane

Let $\vec{n} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ be the normal vector for a plane which contains the point $P_0 = (2, 1, 4)$. Determine if the point $P = (5, 4, 1)$ is contained in this plane.

Solution

By Definition 4.4.2, P is a point in the plane if it satisfies the equation

$$\vec{n} \bullet (\overrightarrow{0P} - \overrightarrow{0P_0}) = 0$$

Given the above $\vec{n}$, P_0, and P, this equation becomes

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \bullet \left(\begin{bmatrix} 5 \\ 4 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \bullet \left(\begin{bmatrix} 3 \\ 3 \\ -3 \end{bmatrix} \right) \\ = 3 + 6 - 9 = 0 \tag{4.4.3}$$

Therefore $P = (5, 4, 1)$ is contained in the plane.

Suppose $\vec{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, $P = (x, y, z)$ and $P_0 = (x_0, y_0, z_0)$.

Then

$$\begin{aligned} \vec{n} \bullet (\overrightarrow{0P} - \overrightarrow{0P_0}) &= 0 \\ \begin{bmatrix} a \\ b \\ c \end{bmatrix} \bullet \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \right) &= 0 \\ \begin{bmatrix} a \\ b \\ c \end{bmatrix} \bullet \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix} &= 0 \\ a(x - x_0) + b(y - y_0) + c(z - z_0) &= 0 \end{aligned} \tag{4.4.4}$$

We can also write this equation as

$$ax + by + cz = ax_0 + by_0 + cz_0$$

Notice that since P_0 is given, $ax_0 + by_0 + cz_0$ is a known scalar, which we can call d. This equation becomes

$$ax + by + cz = d$$

Definition 4.4.5: Scalar Equation of a Plane

Let $\vec{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ be the normal vector for a plane which contains the point $P_0 = (x_0, y_0, z_0)$.Then if $P = (x, y, z)$ is an arbitrary point on the plane, the scalar equation of the plane is given by

$$ax + by + cz = d$$

where $a, b, c, d^{\text{int}}\mathbb{R}$ and $d = ax_0 + by_0 + cz_0$.

Consider the following equation.

Example 4.4.5: Finding the Equation of a Plane

Find an equation of the plane containing $P_0 = (3, -2, 5)$ and orthogonal to $\vec{n} = \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix}$.

Solution

The above vector $\vec{n}$ is the normal vector for this plane. Using Definition 4.4.2, we can determine the vector equation for this plane.

$$\begin{aligned} \vec{n} \bullet (\overrightarrow{0P} - \overrightarrow{0P_0}) &= 0 \\ \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix} \bullet \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} 3 \\ -2 \\ 5 \end{bmatrix} \right) &= 0 \\ \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix} \bullet \begin{bmatrix} x - 3 \\ y + 2 \\ z - 5 \end{bmatrix} &= 0 \end{aligned} \tag{4.4.5}$$

Using Definition 4.4.3, we can determine the scalar equation of the plane.

$$-2x + 4y + 1z = -2(3) + 4(-2) + 1(5) = -9$$

Hence, the vector equation of the plane is

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

and the scalar equation is

$$-2x + 4y + 1z = -9$$

Example 4.4.6: Shortest Distance From a Point to a Plane

Find the shortest distance from the point $P = (3, 2, 3)$ to the plane given by $2x + y + 2z = 2$, and find the point Q on the plane that is closest to P.

Solution

Pick an arbitrary point P_0 on the plane. Then, it follows that

$$\overrightarrow{QP} = proj_{\vec{n}} \overrightarrow{P_0P}$$

and $\|\overrightarrow{QP}\|$ is the shortest distance from P to the plane. Further, the vector $\overrightarrow{0Q} = \overrightarrow{0P} - \overrightarrow{QP}$ gives the necessary point Q.

From the above scalar equation, we have that $\vec{n} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$. Now, choose $P_0 = (1, 0, 0)$ so that $\vec{n} \bullet \overrightarrow{0P} = 2 = d$. Then, $\overrightarrow{P_0P} = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$.

Next, compute $\overrightarrow{QP} = proj_{\vec{n}} \overrightarrow{P_0P}$.

$$\begin{aligned} \overrightarrow{QP} &= proj_{\vec{n}} \overrightarrow{P_0P} \\ &= \left(\frac{\overrightarrow{P_0P} \bullet \vec{n}}{\|\vec{n}\|^2} \right) \vec{n} \\ &= \frac{12}{9} \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \\ &= \frac{4}{3} \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \end{aligned} \tag{4.4.6}$$

Then, $\|\overrightarrow{QP}\| = 4$ so the shortest distance from P to the plane is 4.

Next, to find the point Q on the plane which is closest to P we have

$$\begin{aligned} \overrightarrow{0Q} &= \overrightarrow{0P} - \overrightarrow{QP} \\ &= \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} - \frac{4}{3} \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \end{aligned} \tag{4.4.7}$$

Therefore, $Q = (\frac{1}{3}, \frac{2}{3}, \frac{1}{3})$.

Answers

Proposition $4.4.1$**: Algebraic Description of a Straight Line**

Let $\vec{x}_1, \vec{x}_2 \overset{\text{int}}{\mathbb{R}^n}$. Define $\vec{x}_1 = \vec{a}$ and let $\vec{x}_2 - \vec{x}_1 = \vec{b}$. Since $\vec{b} \neq \vec{0}$, it follows that $\vec{x}_2 \neq \vec{x}_1$. Then $\vec{a} + t\vec{b} = \vec{x}_1 + t\left(\vec{x}_2 - \vec{x}_1\right)$. It follows that $\vec{x} = \vec{a} + t\vec{b}$ is a line containing the two different points X_1 and X_2 whose position vectors are given by $\vec{x}_1$ and $\vec{x}_2$ respectively.

This page titled 4.4: Lines and Planes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

- 4.6: Parametric Lines by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.
- 4.8: Planes in Rⁿ by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.4E: Exercises for Section 4.3

Exercise 4.4E.1

Find the vector equation for the line through $(-7, 6, 0)$ and $(-1, 1, 4)$. Then, find the parametric equations for this line.

Answer available at the end of this section.

Exercise 4.4E.2

Find parametric equations for the line through the point $(7, 7, 1)$ with a direction vector $\vec{d} = \begin{bmatrix} 1 \\ 6 \\ 2 \end{bmatrix}$.

Exercise 4.4E.3

Parametric equations of the line are

$$\begin{aligned} x &= t + 2 \\ y &= 6 - 3t \\ x &= -t = 6 \end{aligned} \tag{4.4E.1}$$

Find a direction vector for the line and a point on the line.

Exercise 4.4E.4

Find the vector equation for the line through the two points $(-5, 5, 1),\ (2, 2, 4)$. Then, find the parametric equations.

Answer available at the end of this section.

Exercise 4.4E.5

The equation of a line in two dimensions is written as $y = x - 5$. Find parametric equations for this line.

Exercise 4.4E.6

Find parametric equations for the line through $(6, 5, -2)$ and $(5, 1, 2)$.

Exercise 4.4E.7

Find the vector equation and parametric equations for the line through the point $(-7, 10, -6)$ with a direction vector $\vec{d} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$.

Exercise $4.4E.8$

Parametric equations of the line are

$$\begin{aligned} x &= 2t + 2 \\ y &= 5 - 4t \backslash \text{nonumver} \\ z &= -t - 3 \end{aligned} \tag{4.4E.2}$$

Find a direction vector for the line and a point on the line, and write the vector equation of the line.

Answer available at the end of this section.

Exercise $4.4E.9$

Find the vector equation and parametric equations for the line through the two points $(4, 10, 0),\ \ (1, -5, -6)$.

Exercise $4.4E.10$

Find the point on the line segment from $P = (-4, 7, 5)$ to $Q = (2, -2, -3)$ which is $\frac{1}{7}$ of the way from P to Q.

Answer available at the end of this section.

Exercise $4.4E.11$

Suppose a triangle in $\mathbb{R}^n$ has vertices at $P_1,\ \ P_2$, and P_3. Consider the lines which are drawn from a vertex to the mid point of the opposite side. Show these three lines intersect in a point and find the coordinates of this point.

Answers

Exercise $4.4E.1$

Vector Equation:

$$\vec{r}(t) = \begin{bmatrix} -7 \\ 6 \\ 0 \end{bmatrix} + t \begin{bmatrix} 6 \\ -5 \\ 4 \end{bmatrix}, \quad t^{\text{int}}\mathbb{R}.$$

Parametric Equations:

$$x = -7 + 6t, \quad y = 6 - 5t, \quad z = 4t, \quad t^{\text{int}}\mathbb{R}.$$

Exercise $4.4E.4$

Vector Equation:

$$\vec{r}(t) = \begin{bmatrix} -5 \\ 5 \\ 1 \end{bmatrix} + t \begin{bmatrix} 7 \\ -3 \\ 3 \end{bmatrix}, \quad t^{\text{int}}\mathbb{R}.$$

Parametric Equations:

$$x = -5 + 7t, \quad y = 5 - 3t, \quad z = 1 + 3t, \quad t^{\text{int}}\mathbb{R}.$$

Exercise $4.4E.8$

Direction Vector: The coefficients of t in the parametric equations give the direction vector:

$$\vec{d} = \begin{bmatrix} 2 \\ -4 \\ -1 \end{bmatrix}.$$

A Point on the Line: Setting $t = 0$ in the parametric equations gives a point on the line:

$$P = (2, 5, -3).$$

Vector Equation: The vector equation of the line is:

$$\vec{r}(t) = \begin{bmatrix} 2 \\ 5 \\ -3 \end{bmatrix} + t \begin{bmatrix} 2 \\ -4 \\ -1 \end{bmatrix}, \quad t^{\text{int}}\mathbb{R}.$$

Exercise $4.4E.10$

The point that is $\frac{1}{7}$ of the way from $P = (-4, 7, 5)$ to $Q = (2, -2, -3)$ is:

$$\left(\frac{-22}{7}, \frac{40}{7}, \frac{27}{7}\right).$$

This page titled 4.4E: Exercises for Section 4.3 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

4.5: Spanning Sets in Rⁿ

Learning Objectives

- Determine the span of a set of vectors.
- Determine if a vector is contained in a specified span.

By generating all linear combinations of a set of vectors one can obtain various subsets of $\mathbb{R}^n$ which we call subspaces. For example what set of vectors in $\mathbb{R}^3$ generate the XY-plane? What is the smallest such set of vectors can you find? The tools of spanning, linear independence and basis are exactly what is needed to answer these and similar questions and are the focus of this section. The following definition is essential.

Definition 4.5.1: 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

$$U \subseteq W$$

We begin this section with a definition.

Definition 4.5.2: 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 4.5.1: Span of Vectors

Describe the span of the vectors $\vec{u} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T$ and Double exponent: use braces to clarify .

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. This concept is explained below in a interactive video format. Feel free to click on it and see the theory in action.

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 4.5.3: Vector in a Span

Let $\vec{u}=\begin{bmatrix}1 & 1 & 0\end{bmatrix}^T$ and Double exponent: use braces to clarify . Show that $\vec{w}=\begin{bmatrix}4 & 5 & 0\end{bmatrix}^T$ 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 $\vec{w}^{\text{int}}\text{span}\{\vec{u},\vec{v}\}$, 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

$$\begin{aligned}a+3b&=4\\a+2b&=5\end{aligned}\tag{4.5.1}$$

We solving this system the usual way, constructing the augmented matrix and row reducing to find the reduced row-echelon form.

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

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 $\text{span}\{\vec{u},\vec{v}\}$.

Example 4.5.4

Let $V=\{\langle 5,-6,-1\rangle,\langle 1,4,5\rangle\}$. Which of the following vectors are in the space spanned by V ?

- $\langle 5,-5,-3\rangle$
- $\langle -1,-4,5\rangle$
- $\langle 2,1,-4\rangle$
- $\langle -5,-1,-1\rangle$
- $\langle -5,2,-3\rangle$

Solution

- **Video Length**: 7 minutes 41 seconds
- **Context**: This video demonstrates how to determine which given vectors lie in the span of a set of vectors by checking if they can be expressed as linear combinations of the spanning set.

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This page titled 4.5: Spanning Sets in R^n is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

- 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.

4.5.E: Exercise for Section 4.4

Exercise $4.5.E.1$

Here are some vectors.

$$\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ 7 \\ -4 \end{bmatrix}, \begin{bmatrix} 5 \\ 7 \\ -10 \end{bmatrix}, \begin{bmatrix} 12 \\ 17 \\ -24 \end{bmatrix}$$

Describe the span of these vectors as the span of as few vectors as possible.

Answer available at the end of this section.

Exercise $4.5.E.2$

Here are some vectors.

$$\begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 12 \\ 29 \\ -24 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ 9 \\ -4 \end{bmatrix}, \begin{bmatrix} 5 \\ 12 \\ -10 \end{bmatrix}.$$

Describe the span of these vectors as the span of as few vectors as possible.

Exercise $4.5.E.3$

Here are some vectors.

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

Describe the span of these vectors as the span of as few vectors as possible.

Exercise $4.5.E.4$

Here are some vectors.

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

Now here is another vector:

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

Is this vector in the span of the first four vectors? If it is, exhibit a linear combination of the first four vectors which equals this vector, using as few vectors as possible in the linear combination.

Answer available at the end of this section.

Exercise $4.5.E.5$

Here are some vectors.

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

Now here is another vector:

$$\begin{bmatrix} 2 \\ -3 \\ -4 \end{bmatrix}$$

Is this vector in the span of the first four vectors? If it is, exhibit a linear combination of the first four vectors which equals this vector, using as few vectors as possible in the linear combination.

Exercise $4.5.E.6$

Here are some vectors.

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

Now here is another vector:

$$\begin{bmatrix} 1 \\ 9 \\ 1 \end{bmatrix}$$

Is this vector in the span of the first four vectors? If it is, exhibit a linear combination of the first four vectors which equals this vector, using as few vectors as possible in the linear combination.

Exercise $4.5.E.7$

Here are some vectors,

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

Now here is another vector:

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

Is this vector in the span of the first four vectors? If it is, exhibit a linear combination of the first four vectors which equals this vector, using as few vectors as possible in the linear combination.

Exercise $4.5.E.8$

Here are some vectors.

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

CC BY

Now here is another vector:

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

Is this vector in the span of the first four vectors? If it is, exhibit a linear combination of the first four vectors which equals this vector, using as few vectors as possible in the linear combination.

Exercise $4.5.E.9$

Here are some vectors.

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

Now here is another vector:

$$\begin{bmatrix} -1 \\ -4 \\ 2 \end{bmatrix}$$

Is this vector in the span of the first four vectors? If it is, exhibit a linear combination of the first four vectors which equals this vector, using as few vectors as possible in the linear combination.

Exercise $4.5.E.10$

Suppose $\{\vec{x}_1, \cdots, \vec{x}_k\}$ is a set of vectors from $\mathbb{R}^n$. Show that $\vec{0}$ is in $\text{span}\{\vec{x}_1, \cdots, \vec{x}_k\}$.

Answer available at the end of this section.

Exercise $4.5.E.11$

Prove that the span of any set of vectors contains the zero vector.

Exercise $4.5.E.12$

Consider the vectors $\vec{v}_1 = \langle 2, 3, -1 \rangle$ and $\vec{v}_2 = \langle 0, 1, 2 \rangle$.

a. Describe $\text{span}(\vec{v}_1, \vec{v}_2)$ geometrically.
b. Express $\text{span}(\vec{v}_1, \vec{v}_2)$ in standard form for the type of object you concluded it is in part a.

Answer

a. Geometric Interpretation: The span of the two given vectors forms a plane through the origin in $\mathbb{R}^3$.
b. Standard Form of the Plane: The equation of the plane is:

$$7x - 4y + 2z = 0.$$

.

Exercise $4.5.E.13$

Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a set of vectors that spans a vector space V. Suppose one of the vectors, say $\vec{v}_i$, is a linear combination of the other vectors in S. Prove that the set $S' = S \setminus \{\vec{v}_i\}$ still spans V, i.e.,

$$\text{span}(S) = \text{span}(S').$$

Note: $S \setminus \{\vec{v}_i\}$ means the set S with the vector $\vec{v}_i\}$ removed.

Exercise $4.5.E.14$

Prove that if a set of vectors contains the zero vector, removing the zero vector does not change the span of the set

Answers

Exercise $4.5.E.1$

The given vectors span the same set as:

$$\text{span}\left\{\begin{bmatrix}1\\1\\-2\end{bmatrix}, \begin{bmatrix}1\\2\\-2\end{bmatrix}\right\}.$$

Exercise $4.5.E.4$

The vector

$$\begin{bmatrix}1\\2\\-1\end{bmatrix} \tag{4.5.E.1}$$

is in the span of the given four vectors. A minimal linear combination expressing this vector is:

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

Exercise $4.5.E.10$

$\sum_{i=1}^{k} 0\vec{x}_k = \vec{0}$

This page titled 4.5.E: Exercise for Section 4.4 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

4.6: Linear Independence

Learning Objectives

- Determine if a set of vectors is linearly independent.
- Evaluate the linear independence of vectors through theoretical analysis and solving systems of equations.

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 $\{\vec{u}_3, \cdots, \vec{u}_k\}$. But sometimes it can be more subtle.

Example 4.6.1: Linearly Dependent Set of Vectors

Consider the vectors

$$\vec{u}_1 = \begin{bmatrix} 0 & 1 & -2 \end{bmatrix}^T, \vec{u}_2 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T, \vec{u}_3 = \begin{bmatrix} -2 & 3 & 2 \end{bmatrix}^T, \text{ and } \vec{u}_4 = \begin{bmatrix} 1 & -2 & 0 \end{bmatrix}^T$$

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 4.6.1: 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 4.6.2: 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 4.6.2: 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.

Example 4.6.3

Consider the vectors: $\vec{v_1} = \begin{bmatrix} 5 & 2 & -3 \end{bmatrix}^T$, $\vec{v_2} = \begin{bmatrix} 14 & 13 & 5 \end{bmatrix}^T$, and $\vec{v_3} = \begin{bmatrix} -3 & 5 & 5 \end{bmatrix}^T$ in $\mathbb{R}^3$. Are these vectors linearly independent?

Solution

- **Video Length**: 8 minutes 8 seconds
- **Context**: Determine if the given set of vectors in $\mathbb{R}^3$ is linearly independent.

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

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 4.6.1: 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 4.6.4: 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

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

The reduced row-echelon form is

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

and so every column is a pivot column and the corresponding system $AX = 0$ only has the trivial solution. Therefore, these vectors are linearly independent and there is no way to obtain one of the vectors as a linear combination of the others.

Consider another example.

Example 4.6.5: Linear Independence

Determine whether the set of vectors given by

$$\left\{\begin{bmatrix}1\\2\\3\\0\end{bmatrix},\begin{bmatrix}2\\1\\0\\1\end{bmatrix},\begin{bmatrix}0\\1\\1\\2\end{bmatrix},\begin{bmatrix}3\\2\\2\\-1\end{bmatrix}\right\}$$

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 $4.6.1$, 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

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

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 $4.6.6$: 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,

$$\begin{aligned} a+2b&=0\\ a+c&=0\\ b-5c&=0 \end{aligned} \tag{4.6.1}$$

This system of three equations in three variables has the unique solution $a=b=c=0$. Therefore, $\{\vec{u}+\vec{v},2\vec{u}+\vec{w},\vec{v}-5\vec{w}\}$ is independent.

The following corollary follows from the fact that if the augmented matrix of a homogeneous system of linear equations has more columns than rows, the system has infinitely many solutions.

Corollary 4.6.1: Linear Dependence in $\mathbb{R}^n$

Let $\{\vec{u}_1,\cdots,\vec{u}_k\}$ be a set of vectors in $\mathbb{R}^n$. If $k>n$, then the set is linearly dependent (i.e. NOT linearly independent).

Proof available at the end of this section.

Example 4.6.7: 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 4.6.1 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 4.6.2: 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 available at the end of this section.

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 4.6.8

Suppose that $\vec{u}, \vec{v}$ and $\vec{w}$ are nonzero vectors in $\mathbb{R}^3$, and that $\{\vec{v}, \vec{w}\}$ is independent. Prove that $\{\vec{u}, \vec{v}, \vec{w}\}$ is independent if and only if $\vec{u} \not\in \text{span}\{\vec{v}, \vec{w}\}$.

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 $\vec{u} \not\in \text{span}\{\vec{v}, \vec{w}\}$, 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 4.6.3: 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.

$$\begin{aligned} CO + \tfrac{1}{2}O_2 &\to CO_2 \\ H_2 + \tfrac{1}{2}O_2 &\to H_2O \\ CH_4 + \tfrac{3}{2}O_2 &\to CO + 2H_2O \\ CH_4 + 2O_2 &\to CO_2 + 2H_2O \end{aligned}$$

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

$$\begin{gathered} CO + 3H_2 - 1H_2O - 1CH_4 = 0 \\ O_2 + 2H_2 - 2H_2O = 0 \\ CO_2 + 4H_2 - 2H_2O - 1CH_4 = 0 \end{gathered}$$

In terms of the original notation, these are the reactions

$$\begin{gathered} CO + 3H_2 \to H_2O + CH_4 \\ O_2 + 2H_2 \to 2H_2O \\ CO_2 + 4H_2 \to 2H_2O + CH_4 \end{gathered}$$

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.

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.

Answers

Corollary $4.6.1$**: Linear Dependence in** $\mathbb{R}''$

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

Theorem $4.6.2$**: Unique Linear Combination**

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 $\vec{x}^{\text{int}}\text{span}(U)$ such that

$$\begin{aligned} \vec{x} &= s_1\vec{u}_1 + s_2\vec{u}_2 + \cdots + s_k\vec{u}_k, \text{ for some } s_1, s_2, \ldots, s_k^{\text{int}}\mathbb{R}, \text{ and} \\ \vec{x} &= t_1\vec{u}_1 + t_2\vec{u}_2 + \cdots + t_k\vec{u}_k, \text{ for some } t_1, t_2, \ldots, t_k^{\text{int}}\mathbb{R}. \end{aligned} \tag{4.6.2}$$

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 $\vec{x}^{\text{int}}\text{span}(U)$ can be written uniquely as a linear combination of vectors of U.

This page titled 4.6: Linear Independence is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

- 4.10: Spanning, Linear Independence and Basis in Rⁿ by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.6.E: Exercise for Section 4.5

Exercise 4.6.E.1

Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.

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

Answer available at the end of this section.

Exercise 4.6.E.2

Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.

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

Exercise 4.6.E.3

Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.

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

Exercise 4.6.E.4

Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.

$$\begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \\ 34 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 7 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 8 \\ 1 \end{bmatrix}$$

Answer available at the end of this section.

Exercise 4.6.E.5

Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others.

$$\begin{bmatrix} 1 \\ 3 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ -10 \\ 3 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 0 \\ 1 \end{bmatrix}$$

Exercise 4.6.E.6

Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.

$$\begin{bmatrix} 1 \\ 3 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ -5 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ -4 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 10 \\ -14 \\ 1 \end{bmatrix}$$

Exercise 4.6.E.7

Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.

$$\begin{bmatrix} 1 \\ 0 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 8 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 7 \\ 34 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 7 \\ 1 \end{bmatrix}$$

Exercise 4.6.E.8

Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.

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

Exercise 4.6.E.9

Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others.

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

Answer available at the end of this section.

Exercise 4.6.E.10

Are the following vectors linearly independent? If they are, explain why and if they are not, exhibit one of them as a linear combination of the others. Also give a linearly independent set of vectors which has the same span as the given vectors.

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

Exercise $4.6.E.11$

Suppose $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a linearly dependent set of vectors in $\mathbb{R}^n$. Prove that at least one of these vectors can be written as a linear combination of the others.

Answer available at the end of this section.

Exercise $4.6.E.12$

Let $B = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n\}$ be a linearly independent set. Prove that every vector in $\text{span}(B)$ can be written uniquely as a linear combination of the vectors in B.

Exercise $4.6.E.13$

Prove that any set containing the zero vector is a linearly dependent set.

Answer available at the end of this section.

Exercise $4.6.E.14$

Suppose $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are linearly independent vectors in $\mathbb{R}^3$. Prove that $\vec{v}_1 + \vec{v}_2, \vec{v}_2 + \vec{v}_3, \vec{v}_3 + \vec{v}_1$ are also linearly independent

Exercise $4.6.E.15$

Let A be an $n \times n$ matrix. Prove that the column vectors of A are linearly independent if and only if $\det(A) \neq 0$.

Hint available at the end of this section.

Exercise $4.6.E.16$

Suppose the set $B = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n\}$ of vectors spans $\mathbb{R}^n$. Suppose $\vec{v}$ is any vector in $\mathbb{R}^n$. Prove that $B \cup \{\vec{v}\}$ is a linearly dependent set

Exercise $4.6.E.17$

Suppose $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is a linearly independent set of vectors. Prove that any subset of S is also linearly independent

Answers

Exercise $4.6.E.1$

The given vectors are linearly dependent. We express one of them as a linear combination of the others:

$$\vec{v}_4 = -2\vec{v}_1 + 3\vec{v}_2 + 1\vec{v}_3.$$

A linearly independent set that spans the same space is:

$$\left\{\begin{bmatrix}1\\3\\-1\\1\end{bmatrix},\begin{bmatrix}1\\4\\-1\\1\end{bmatrix},\begin{bmatrix}1\\4\\0\\1\end{bmatrix}\right\}.$$

Exercise $4.6.E.4$

The given vectors are linearly dependent. We express one of them as a linear combination of the others:

$$\vec{v}_4 = -\frac{13.43}{11.43}\vec{v}_3 + \text{other terms.}$$

A linearly independent set that spans the same space is:

$$\left\{\begin{bmatrix}1\\-1\\3\\1\end{bmatrix},\begin{bmatrix}1\\6\\34\\1\end{bmatrix},\begin{bmatrix}1\\0\\7\\1\end{bmatrix}\right\}.$$

Exercise $4.6.E.9$

The given vectors are linearly independent because the only solution to the homogeneous system

$$a\begin{bmatrix}1\\2\\2\\-4\end{bmatrix}+b\begin{bmatrix}3\\4\\1\\-4\end{bmatrix}+c\begin{bmatrix}0\\-1\\0\\4\end{bmatrix}+d\begin{bmatrix}0\\-1\\-2\\5\end{bmatrix}=\begin{bmatrix}0\\0\\0\\0\end{bmatrix}$$

is the trivial solution $a=b=c=d=0$. Therefore, the vectors are linearly independent.

Exercise $4.6.E.11$

We are given that the set $S=\{\vec{v}_1,\vec{v}_2,\dots,\vec{v}_n\}$ is \textbf{linearly dependent} in $\mathbb{R}^n$. By definition, this means that there exist scalars $a_1,a_2,\dots,a_n$, not all zero, such that:

$$a_1\vec{v}_1+a_2\vec{v}_2+\cdots+a_n\vec{v}_n=\vec{0}.$$

Since not all of the a_i are zero, at least one of them is nonzero. Without loss of generality, assume that $a_k\neq 0$ for some k, meaning the term $a_k\vec{v}_k$ is present in the equation. We can solve for $\vec{v}_k$ in terms of the other vectors:

$$a_k\vec{v}_k=-(a_1\vec{v}_1+a_2\vec{v}_2+\cdots+a_{k-1}\vec{v}_{k-1}+a_{k+1}\vec{v}_{k+1}+\cdots+a_n\vec{v}_n).$$

Since $a_k\neq 0$, we can divide by a_k to express $\vec{v}_k$ as a linear combination of the other vectors:

$$\vec{v}_k=-\frac{a_1}{a_k}\vec{v}_1-\frac{a_2}{a_k}\vec{v}_2-\cdots-\frac{a_{k-1}}{a_k}\vec{v}_{k-1}-\frac{a_{k+1}}{a_k}\vec{v}_{k+1}-\cdots-\frac{a_n}{a_k}\vec{v}_n.$$

Thus, we have shown that at least one vector in S can be written as a linear combination of the others.

Exercise $4.6.E.13$

Let $S=\{\vec{v}_1,\vec{v}_2,\dots,\vec{v}_k,\vec{0}\}$ be a set of vectors in $\mathbb{R}^n$ that contains the zero vector $\vec{0}$. By the definition of linear dependence, we need to check if there exists a nontrivial linear combination of the vectors in S that equals the zero vector:

$$a_1\vec{v}_1+a_2\vec{v}_2+\cdots+a_k\vec{v}_k+a_{k+1}\vec{0}=\vec{0}$$

for some scalars $a_1,a_2,\dots,a_k,a_{k+1}$, not all zero. Consider setting $a_{k+1}=1$ and all other coefficients $a_1,a_2,\dots,a_k$ equal to zero:

$$0\cdot\vec{v}_1+0\cdot\vec{v}_2+\cdots+0\cdot\vec{v}_k+1\cdot\vec{0}=\vec{0}.$$

Since this is a nontrivial linear combination (not all coefficients are zero), the set S is linearly dependent.

Exercise $4.6.E.15$

Since this is an "if and only if" proof, you must prove each direction.

Let A be an $n\times n$ matrix with column vectors $\vec{a}_1,\vec{a}_2,\dots,\vec{a}_n$. We need to prove that the column vectors of A are linearly independent if and only if $\det(A)\neq 0$.

($\Leftarrow$) Suppose $\det(A)\neq 0$. [We must show that the column vectors of A are linearly independent.] Consider the homogeneous system of equations:

$$A\vec{x}=\vec{0}$$

where $\vec{x}=\langle x_1,x_2,\dots,x_n\rangle^T$ is an $n\times 1$ vector of unknowns. (What does that fact that $\det(A)=0$ say about the solution to $A\vec{x}=\vec{0}$? Finish the argument.)

($\Rightarrow$) Suppose the column vectors of A are linearly independent. [We must show that $\det(A)\neq 0$.] If the column vectors of A are linearly independent, then the only solution to the homogeneous system $A\vec{x}=\vec{0}$ is the trivial solution $\vec{x}=\vec{0}$ since each column of A is a pivot column. (Finish the proof from here.)

This page titled 4.6.E: Exercise for Section 4.5 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

4.7: 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.

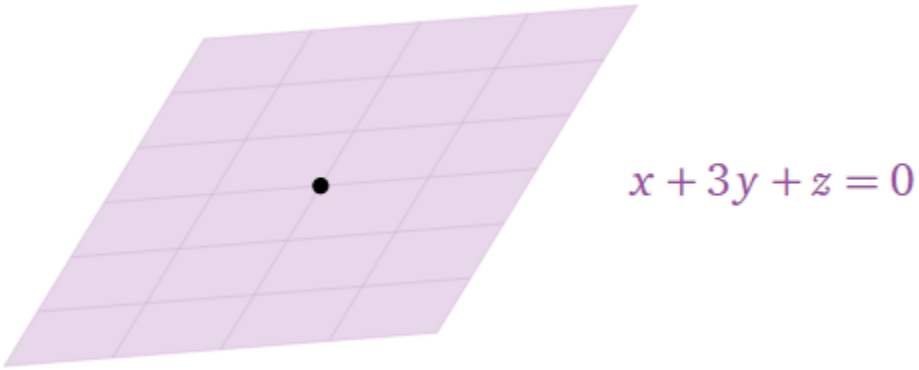


Figure 4.7.1

(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 4.7.1: Subset

A ***subset*** of $\mathbb{R}^n$ is any collection of points of $\mathbb{R}^n$.

For instance, the unit circle

$$C = \{(x, y) \text{ in } \mathbb{R}^2 \mid x^2 + y^2 = 1\}$$

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 4.7.2: 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 4.7.1 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 4.7.1

The set $\mathbb{R}^n$ is a subspace of itself: indeed, it contains zero, and is closed under addition and scalar multiplication.

Example 4.7.2

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 4.7.3: A line through the origin

A line L through the origin is a subspace.

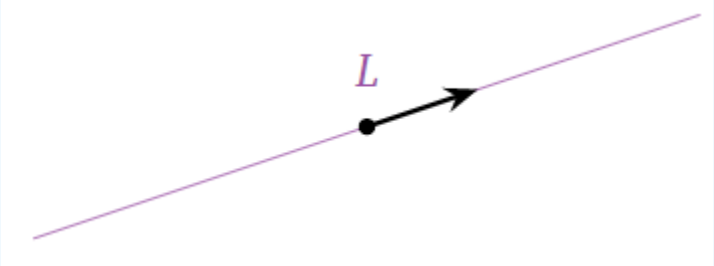


Figure 4.7.2

Indeed, L contains zero, and is easily seen to be closed under addition and scalar multiplication.

Example 4.7.4: A plane through the origin

A plane P through the origin is a subspace.

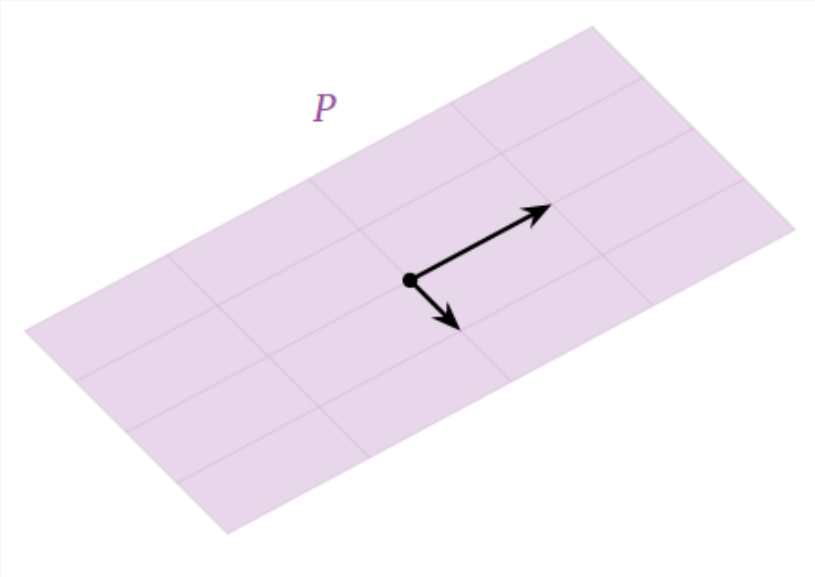


Figure 4.7.3

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 4.7.5: 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.

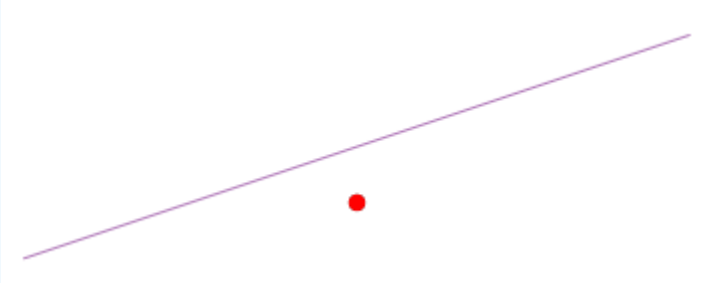


Figure 4.7.4

Example 4.7.6: 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.

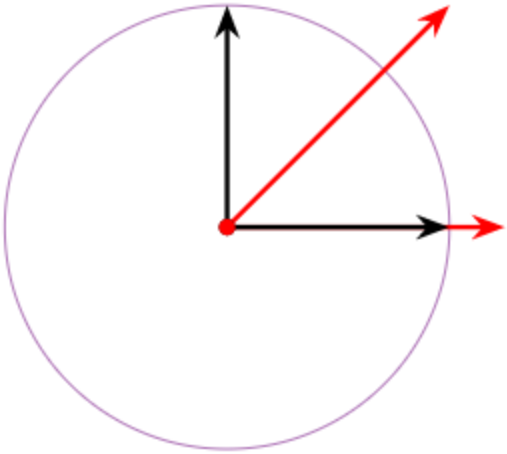


Figure 4.7.5

Example 4.7.7: 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).

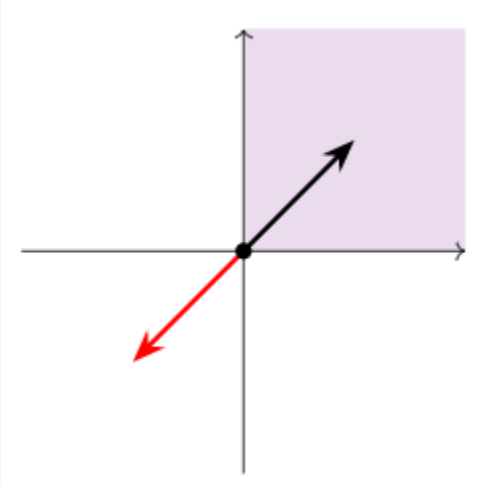


Figure 4.7.6

Example 4.7.8: 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.

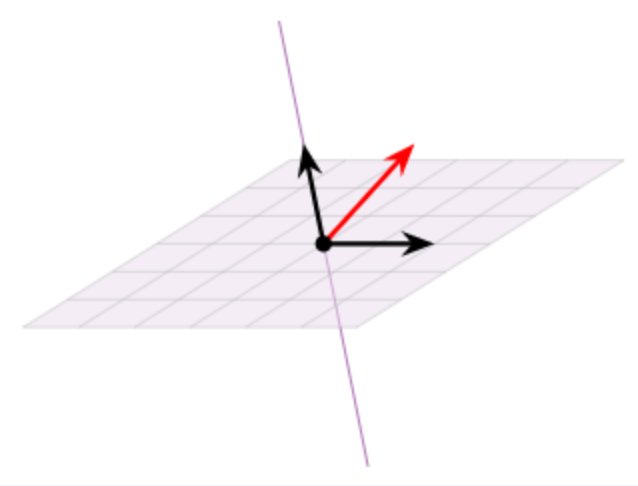


Figure 4.7.7

Note 4.7.1: Subsets versus Subspaces

A subset of $\mathbb{R}^n$ is any collection of vectors whatsoever. For instance, the unit circle

$$C = \left\{(x, y) \text{ in } \mathbb{R}^2 \mid x^2 + y^2 = 1\right\}$$

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 4.7.3, below.

Example 4.7.9: 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 $\begin{pmatrix} a \\ b \end{pmatrix}$ 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, $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 1/2 \\ 1/3 \end{pmatrix}$ are in V, but $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ is not because $2 \cdot 2 \neq 3 \cdot 3$.

Let us check the first property. The subset V does contain the zero vector $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, 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 $u = \begin{pmatrix} a \\ b \end{pmatrix}$ and $v = \begin{pmatrix} c \\ d \end{pmatrix}$ 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 $\begin{pmatrix} a+c \\ b+d \end{pmatrix}$ 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 $v = \begin{pmatrix} a \\ b \end{pmatrix}$ 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 $\begin{pmatrix} ca \\ cb \end{pmatrix}$ 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 $2/3$.

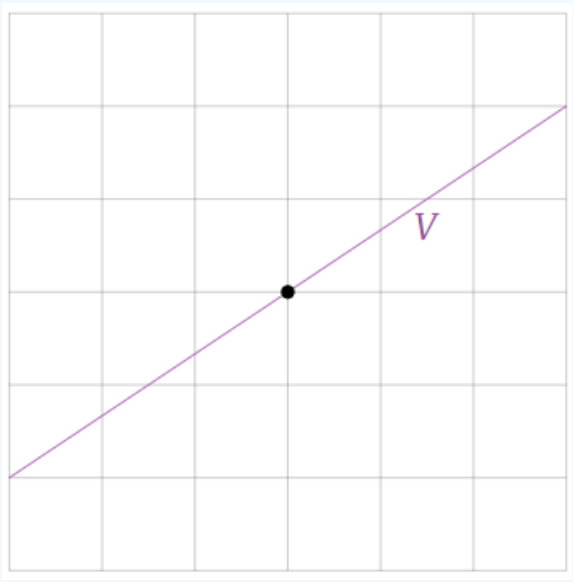


Figure 4.7.8

Example 4.7.10: 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.

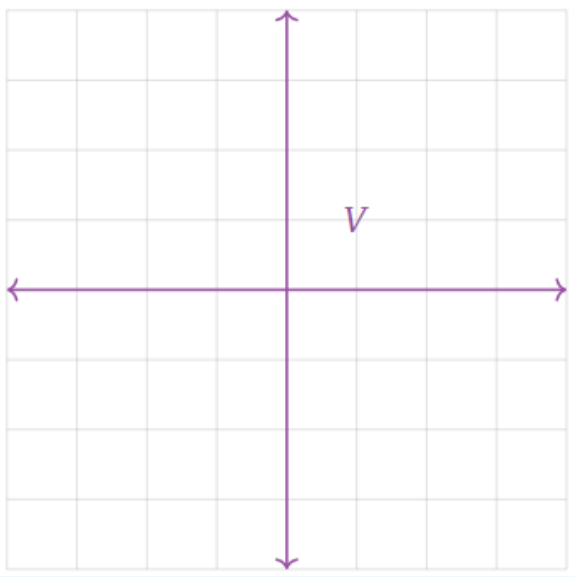


Figure 4.7.9

Common Types of Subspaces

Theorem 4.7.1: 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 available at the end of this section.

If $V = \text{Span}\{v_1, v_2, \dots, v_p\}$, we say that V is the subspace ***spanned by*** or ***generated by*** the vectors $v_1, v_2, \dots, v_p$. We call $\{v_1, v_2, \dots, v_p\}$ a ***spanning set*** for V.

Any matrix naturally gives rise to *two* subspaces.

Definition 4.7.3: 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 4.7.1. 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 4.7.11

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 $\mathbb{R}^3$.

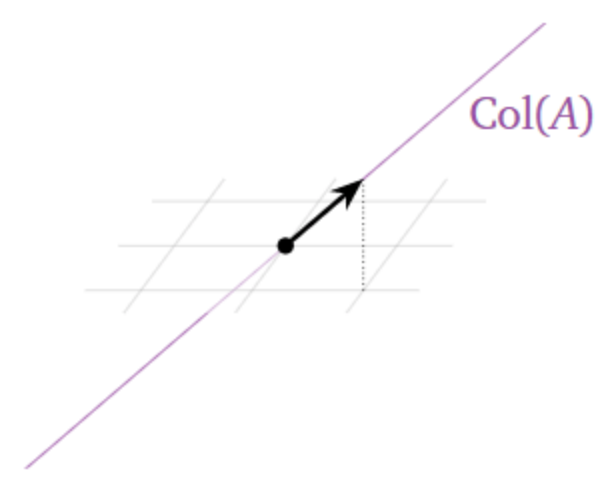


Figure 4.7.10

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 $\mathbb{R}^2$.

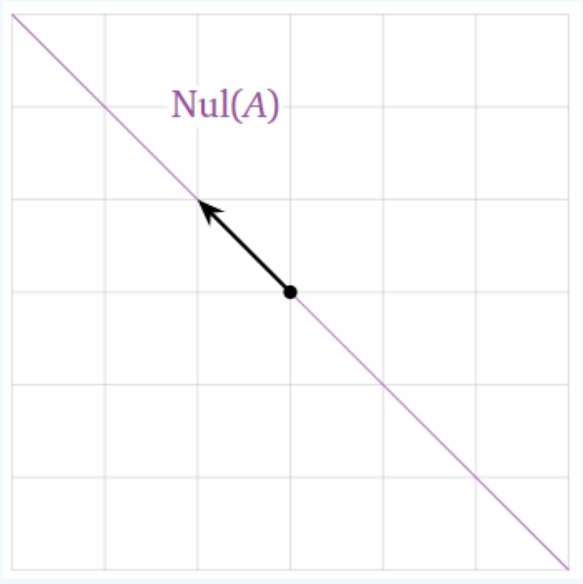


Figure 4.7.11

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 4.7.2: 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 4.7.12: 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,

$$\text{Nul}(A) = \text{Span}\left\{\begin{pmatrix}1\\2\\1\\0\end{pmatrix}, \begin{pmatrix}-2\\3\\0\\1\end{pmatrix}\right\}.$$

Example 4.7.13: 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,

$$\text{Nul}(A) = \{0\} = \text{Span}\{0\}.$$

It is natural to define $\text{Span}\{\} = \{0\}$, 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 4.7.3

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 4.7.14

Let

$$V = \left\{\begin{pmatrix}a\\b\end{pmatrix} \text{ in } \mathbb{R}^2 | 2a = 3b\right\}$$

be the subset of a previous Example 4.7.9. 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 $\begin{pmatrix} 2 & -3 \end{pmatrix}$ is $\begin{pmatrix} 1 & -3/2 \end{pmatrix}$, so the parametric form of V is $x = 3/2y$, so the parametric vector form is $\begin{pmatrix} x \\ y \end{pmatrix} = y\begin{pmatrix} 3/2 \\ 1 \end{pmatrix}$, and hence $\left\{\begin{pmatrix} 3/2 \\ 1 \end{pmatrix}\right\}$ spans V.

Example 4.7.15

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

$$V = \text{Col}\begin{pmatrix} 2 & 1 \\ 1 & -1 \\ 3 & -2 \end{pmatrix}.$$

Answers

Theorem 4.7.1: Spans are Subspaces and Subspaces are Spans

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.

This page titled 4.7: Subspaces is shared under a GNU FDL license and was authored, remixed, and/or curated by Dan Margalit and Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- 2.6: Subspaces by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

4.8: Subspaces and Bases

Learning Objectives

- Define and identify subspaces within $\mathbb{R}^n$.
- Analyze and verify if a set of vectors forms a basis for a given subspace.
- Determine the dimension of a subspace and extend or shrink bases as needed.

The goal of this section is to develop an understanding of a subspace of $\mathbb{R}^n$.

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 $\mathbb{R}^3$.

Definition 4.8.1: 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.

Sometimes, instead of using the definition of subspace, it's easier to use the following test, especially the first part. Since a subspace must be closed under all possible linear combinations, it must contain the trivial linear combination, $\vec{0}$. Hence, if a set doesn't contain $\vec{0}$, it cannot possibly be a subspace.

Theorem 4.8.1: 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 $\vec{u},\vec{w}^{\,\text{int}}V$, $\vec{u}+\vec{w}^{\,\text{int}}V$;
3. V is closed under scalar multiplication, i.e., for all $\vec{u}^{\,\text{int}}V$ and $k^{\,\text{int}}\mathbb{R}$, $k\vec{u}^{\,\text{int}}V$.

Consider the following example of a line in $\mathbb{R}^3$.

Example 4.8.1: 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 $\begin{bmatrix}x\\y\\z\end{bmatrix}=t\begin{bmatrix}-5\\1\\-4\end{bmatrix}, t^{\,\text{int}}\mathbb{R}$,

so

$$L=\left\{t\vec{d}\mid t^{\,\text{int}}\mathbb{R}\right\}.$$

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: $\vec{0}_3^{\,\text{int}}L$ since $0\vec{d}=\vec{0}_3$.
- Suppose $\vec{u},\vec{v}^{\,\text{int}}L$. Then by definition, $\vec{u}=s\vec{d}$ and $\vec{v}=t\vec{d}$, for some $s,t^{\,\text{int}}\mathbb{R}$. Thus

$$\vec{u}+\vec{v}=s\vec{d}+t\vec{d}=(s+t)\vec{d}.$$

Since $s+t^{\text{int}}\mathbb{R}$, $\vec{u}+\vec{v}^{\text{int}}L$; i.e., L is closed under addition.

- Suppose $\vec{u}^{\text{int}}L$ and $k^{\text{int}}\mathbb{R}$ (k is a scalar). Then $\vec{u}=t\vec{d}$, for some $t^{\text{int}}\mathbb{R}$, so

$$k\vec{u}=k(t\vec{d})=(kt)\vec{d}.$$

Since $kt^{\text{int}}\mathbb{R}$, $k\vec{u}^{\text{int}}L$; i.e., L is closed under scalar multiplication.

Since L satisfies all conditions of the subspace test, it follows that L is a subspace.

Theorem 4.8.2: 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 $\{\vec{u}_1,\cdots,\vec{u}_k\}^{\text{int}}W$. 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 available at the end of this section.

Since the vectors $\vec{u}_i$ we constructed in the proof above are not in the span of the previous vectors (by definition), they must be linearly independent and thus we obtain the following corollary.

Corollary 4.8.1: 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 4.8.2: 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 4.8.3: 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 any two bases of a subspace must be of the same size. To show this, we will need the the following fundamental result, called the Exchange Theorem.

Theorem 4.8.3: 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 available at the end of this section.

We are now ready to show that any two bases are of the same size.

Theorem 4.8.4: 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 available at the end of this section.

The following definition can now be stated.

Definition 4.8.4: Dimension of a Subspace

Let V be a subspace of $\mathbb{R}^n$. Then the **dimension** of V, written $\dim(V)$ is defined to be the number of vectors in a basis.

The next result follows.

Corollary 4.8.2: Dimension of $\mathbb{R}^n$

The dimension of $\mathbb{R}^n$ is n.

Proof available at the end of this section.

Consider the following example.

Example 4.8.2: Basis of Subspace

Let

$$V = \left\{ \left[\begin{array}{c} a \\ b \\ c \\ d \end{array} \right]^{\text{int}} \mathbb{R}^4 \ : \ a - b = d - c \right\}.$$

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

$$V = \left\{ \left[\begin{array}{c} b - c + d \\ b \\ c \\ d \end{array} \right] \ : \ b, c, d^{\text{int}}\mathbb{R} \right\} = \left\{ b \left[\begin{array}{c} 1 \\ 1 \\ 0 \\ 0 \end{array} \right] + c \left[\begin{array}{c} -1 \\ 0 \\ 1 \\ 0 \end{array} \right] + d \left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 1 \end{array} \right] \ : \ b, c, d^{\text{int}}\mathbb{R} \right\}$$

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 4.8.3: 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 available at the end of this section.

The next theorem follows from the above claim.

Theorem 4.8.5: Existence of Basis

Let V be a subspace of $\mathbb{R}^n$. Then there exists a basis of V with $\dim(V) \leq n$.

Consider Corollary 4.8.4 together with Theorem 4.8.8. 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 4.8.3: Extending an Independent Set

Consider the set U given by

$$U = \left\{ \left[\begin{array}{c} a \\ b \\ c \\ d \end{array} \right]^{\text{int}}_{\mathbb{R}^4} \left| \begin{array}{l} \\ \\ \\ \\ \end{array} \right. a - b = d - c \right\}$$

Then U is a subspace of $\mathbb{R}^4$ and $\dim(U) = 3$.

Then

$$S = \left\{ \left[\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \end{array} \right], \left[\begin{array}{c} 2 \\ 3 \\ 3 \\ 2 \end{array} \right] \right\},$$

is an independent subset of U. Show that S can be extended to a basis of U.

Solution

To extend S to a basis of U, find a vector in U that is **not** in $\text{span}(S)$.

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

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

Therefore, S can be extended to the following basis of U:

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

Next we consider the case of removing vectors from a spanning set to result in a basis.

Theorem 4.8.6: Finding a Basis from a Span

Let W be a subspace. Also suppose that $W = span\left\{\vec{w}_1, \cdots, \vec{w}_m\right\}$. Then there exists a subset of $\left\{\vec{w}_1, \cdots, \vec{w}_m\right\}$ which is a basis for W.

Proof available at the end of this section.

The following example illustrates how to carry out this shrinking process which will obtain a subset of a span of vectors which is linearly independent.

Example 4.8.4: 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 4.8.7: Subset of a Subspace

Let V and W be subspaces of $\mathbb{R}^n$, and suppose that $W \subseteq V$. Then $\dim(W) \leq \dim(V)$ with equality when $W = V$.

Theorem 4.8.8: Extending a Basis

Let W be any non-zero subspace $\mathbb{R}^n$ and let $W \subseteq V$ where V is also a subspace of $\mathbb{R}^n$. Then every basis of W can be extended to a basis for V.

The proof is left as an exercise but proceeds as follows. Begin with a basis for $W, \{\vec{w}_1, \cdots, \vec{w}_s\}$ and add in vectors from V until you obtain a basis for V. Not that the process will stop because the dimension of V is no more than n.

Consider the following example.

Example 4.8.5: 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

$$\begin{bmatrix}1&0&1&0&0&0\\0&1&0&1&0&0\\1&0&0&0&1&0\\1&1&0&0&0&1\end{bmatrix} \tag{4.8.1}$$

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

$$\begin{bmatrix}1&0&0&0&1&0\\0&1&0&0&-1&1\\0&0&1&0&-1&0\\0&0&0&1&1&-1\end{bmatrix} \tag{4.8.2}$$

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 (4.8.1) 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 4.8.6: 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.

Answers

Theorem 4.8.2**: Subspaces are Spans**

We first show that if V is a subspace, then it can be written as $V = \operatorname{span}\{\vec{u}_1, \cdots, \vec{u}_k\}$. Pick a vector $\vec{u}_1$ in V. If $V = \operatorname{span}\{\vec{u}_1\}$, then you have found your list of vectors and are done. If $V \neq \operatorname{span}\{\vec{u}_1\}$, then there exists $\vec{u}_2$ a vector of V which is not in $\operatorname{span}\{\vec{u}_1\}$. Consider $\operatorname{span}\{\vec{u}_1, \vec{u}_2\}$. If $V = \operatorname{span}\{\vec{u}_1, \vec{u}_2\}$, we are done. Otherwise, pick $\vec{u}_3$ not in $\operatorname{span}\{\vec{u}_1, \vec{u}_2\}$. Continue this way. Note that since V is a subspace, these spans are each contained in V. The process must stop with $\vec{u}_k$ for some $k \leq n$ by Corollary 4.8.1, and thus $V = \operatorname{span}\{\vec{u}_1, \cdots, \vec{u}_k\}$.

Now suppose $V = \operatorname{span}\{\vec{u}_1, \cdots, \vec{u}_k\}$, 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 $\operatorname{span}\{\vec{u}_1, \cdots, \vec{u}_k\}$ and is therefore contained in V. This shows that $\operatorname{span}\{\vec{u}_1, \cdots, \vec{u}_k\}$ 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 $a_1\vec{u}_1 + a_2\vec{u}_2 + \cdots + a_k\vec{u}_k \in W$.

Theorem 4.8.3**: Exchange Theorem**

Since each $\vec{u}_j$ is in $\operatorname{span}\{\vec{v}_1, \cdots, \vec{v}_s\}$, 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

Theorem 4.8.4**: Bases of $\mathbb{R}^n$ are of the Same Size**

This follows from Theorem 4.8.3. 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 $r \geq s$.

Corollary 4.8.2**: Dimension of $\mathbb{R}^n$**

You only need to exhibit a basis for $\mathbb{R}^n$ which has n vectors. Such a basis is the standard basis $\{\vec{e}_1, \cdots, \vec{e}_n\}$.

Corollary 4.8.3**: Linearly Independent and Spanning Sets in $\mathbb{R}^n$**

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

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 4.8.3.

Theorem 4.8.6**: Finding a Basis from a Span**

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 $\{\vec{w}_1, \cdots, \vec{w}_m\}$.

This page titled 4.8: Subspaces and Bases is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

- 4.10: Spanning, Linear Independence and Basis in Rⁿ by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.8.E: Exercise for Section 4.6

Exercise 4.8.E.1

Let $H = \text{span}\left\{\begin{bmatrix}2\\1\\1\\1\end{bmatrix}, \begin{bmatrix}-1\\0\\-1\\-1\end{bmatrix}, \begin{bmatrix}5\\2\\3\\3\end{bmatrix}, \begin{bmatrix}-1\\1\\-2\\-2\end{bmatrix}\right\}$. Find the dimension of H and determine a basis.

Exercise 4.8.E.2

Let $H = \text{span}\left\{\begin{bmatrix}0\\1\\1\\-1\end{bmatrix}, \begin{bmatrix}-1\\-1\\-2\\2\end{bmatrix}, \begin{bmatrix}2\\3\\5\\-5\end{bmatrix}, \begin{bmatrix}0\\1\\2\\-2\end{bmatrix}\right\}$. Find the dimension of H and determine a basis.

Exercise 4.8.E.3

Let $H = \text{span}\left\{\begin{bmatrix}-2\\1\\1\\-3\end{bmatrix}, \begin{bmatrix}-9\\4\\3\\-9\end{bmatrix}, \begin{bmatrix}-33\\15\\12\\-36\end{bmatrix}, \begin{bmatrix}-22\\10\\8\\-24\end{bmatrix}\right\}$. Find the dimension of H and determine a basis.

Exercise 4.8.E.4

Let $H = \text{span}\left\{\begin{bmatrix}-1\\1\\-1\\-2\end{bmatrix}, \begin{bmatrix}-4\\3\\-2\\-4\end{bmatrix}, \begin{bmatrix}-3\\2\\-1\\-2\end{bmatrix}, \begin{bmatrix}-1\\1\\-2\\-4\end{bmatrix}, \begin{bmatrix}-7\\5\\-3\\-6\end{bmatrix}\right\}$. Find the dimension of H and determine a basis.

Exercise 4.8.E.5

Let $H = \text{span}\left\{\begin{bmatrix}2\\3\\2\\1\end{bmatrix}, \begin{bmatrix}8\\15\\6\\3\end{bmatrix}, \begin{bmatrix}3\\6\\2\\1\end{bmatrix}, \begin{bmatrix}4\\6\\6\\3\end{bmatrix}, \begin{bmatrix}8\\15\\6\\3\end{bmatrix}\right\}$. Find the dimension of H and determine a basis.

Exercise 4.8.E.6

Let $H = \text{span}\left\{\begin{bmatrix}0\\2\\0\\-1\end{bmatrix}, \begin{bmatrix}-1\\6\\0\\-2\end{bmatrix}, \begin{bmatrix}-2\\16\\0\\-6\end{bmatrix}, \begin{bmatrix}-3\\22\\0\\-8\end{bmatrix}\right\}$. Find the dimension of H and determine a basis.

Exercise 4.8.E.7

Let $H = \text{span}\left\{\begin{bmatrix}5\\1\\1\\4\end{bmatrix}, \begin{bmatrix}14\\3\\2\\8\end{bmatrix}, \begin{bmatrix}38\\8\\6\\24\end{bmatrix}, \begin{bmatrix}47\\10\\7\\28\end{bmatrix}, \begin{bmatrix}10\\2\\3\\12\end{bmatrix}\right\}$. Find the dimension of H and determine a basis.

Exercise 4.8.E.8

Let $H = \text{span}\left\{\begin{bmatrix}6\\1\\1\\5\end{bmatrix}, \begin{bmatrix}17\\3\\2\\10\end{bmatrix}, \begin{bmatrix}52\\9\\7\\35\end{bmatrix}, \begin{bmatrix}18\\3\\4\\20\end{bmatrix}\right\}$. Find the dimension of H and determine a basis.

Exercise 4.8.E.9

Let $M = \left\{\vec{u} = \begin{bmatrix}u_1\\u_2\\u_3\\u_4\end{bmatrix}^{\text{int}} \mathbb{R}^4 : \sin(u_1) = 1\right\}$. Is M a subspace? Explain.

Answer available at the end of this section.

Exercise 4.8.E.10

Let $M = \left\{\vec{u} = \begin{bmatrix}u_1\\u_2\\u_3\\u_4\end{bmatrix}^{\text{int}} \mathbb{R}^4 : ||u_1|| \leq 4\right\}$. Is M a subspace? Explain.

Answer available at the end of this section.

Exercise 4.8.E.11

Let $M = \left\{\vec{u} = \begin{bmatrix}u_1\\u_2\\u_3\\u_4\end{bmatrix}^{\text{int}} \mathbb{R}^4 : u_1 \geq 0 \text{ for each } i = 1, 2, 3, 4\right\}$. Is M a subspace? Explain.

Answer available at the end of this section.

Exercise 4.8.E.12

Let $\vec{w}$, $\vec{w}_1$ be given vectors in $\mathbb{R}^4$ and define

$$M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}^{\text{int}} \mathbb{R}^4 : \vec{w} \bullet \vec{u} = 0 \text{ and } \vec{w}_1 \bullet \vec{u} = 0 \right\}.$$

Is M a subspace? Explain.

Answer available at the end of this section.

Exercise $4.8.E.13$

Let $\vec{w}^{\text{int}}\mathbb{R}^4$ and let $M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}^{\text{int}} \mathbb{R}^4 : \vec{w} \bullet \vec{u} = 0 \right\}$. Is M a subspace? Explain.

Answer available at the end of this section.

Exercise $4.8.E.14$

Let $M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}^{\text{int}} \mathbb{R}^4 : u_3 \geq u_1 \right\}$. Is M a subspace? Explain.

Answer available at the end of this section.

Exercise $4.8.E.15$

Let $M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}^{\text{int}} \mathbb{R}^4 : u_3 = u_1 = 0 \right\}$. Is M a subspace? Explain.

Answer available at the end of this section.

Exercise $4.8.E.16$

Consider the set of vectors S given by

$$S = \left\{ \begin{bmatrix} 4u + v - 5w \\ 12u + 6v - 6w \\ 4u + 4v + 4w \end{bmatrix} : u, v, w^{\text{int}}\mathbb{R} \right\}.$$

Is S a subspace of $\mathbb{R}^3$? If so, explain why, give a basis for the subspace and find its dimension.

Exercise $4.8.E.17$

Consider the set of vectors S given by

$$S = \left\{ \begin{bmatrix} 2u+6v+7w \\ -3u-9v-12w \\ 2u+6v+6w \\ u+3v+3w \end{bmatrix} : u, v, w \in \mathbb{R} \right\}.$$

Is S a subspace of $\mathbb{R}^4$? If so, explain why, give a basis for the subspace and find its dimension.

Exercise 4.8.E.18

Consider the set of vectors S given by

$$S = \left\{ \begin{bmatrix} 2u+v \\ 6v-3u+3w \\ 3v-6u+3w \end{bmatrix} : u, v, w \in \mathbb{R} \right\}.$$

Is this set of vectors a subspace of $\mathbb{R}^3$? If so, explain why, give a basis for the subspace and find its dimension.

Exercise 4.8.E.19

Consider the vectors of the form

$$\left\{ \begin{bmatrix} 2u+v+7w \\ u-2v+w \\ -6v-6w \end{bmatrix} : u, v, w \in \mathbb{R} \right\}.$$

Is this set of vectors a subspace of $\mathbb{R}^3$? If so, explain why, give a basis for the subspace and find its dimension.

Exercise 4.8.E.20

Consider the vectors of the form

$$\left\{ \begin{bmatrix} 3u+v+11w \\ 18u+6v+66w \\ 28u+8v+100w \end{bmatrix} : u, v, w \in \mathbb{R} \right\}.$$

Is this set of vectors a subspace of $\mathbb{R}^3$? If so, explain why, give a basis for the subspace and find its dimension.

Exercise 4.8.E.21

Consider the vectors of the form

$$\left\{ \begin{bmatrix} 3u+v \\ 2w-4u \\ 2w-2v-8u \end{bmatrix} : u, v, w \in \mathbb{R} \right\}.$$

Is this set of vectors a subspace of $\mathbb{R}^3$? If so, explain why, give a basis for the subspace and find its dimension.

Exercise 4.8.E.22

Consider the set of vectors S given by

$$\left\{ \begin{bmatrix} u+v+w \\ 2u+2v+4w \\ u+v+w \\ 0 \end{bmatrix} : u, v, w \in \mathbb{R} \right\}.$$

Is S is a subspace of $\mathbb{R}^4$? If so, explain why, give a basis for the subspace and find its dimension.

Exercise $4.8.E.23$

Consider the set of vectors S given by

$$\left\{ \begin{bmatrix} v \\ -3u-3w \\ 8u-4v+4w \end{bmatrix} : u, v, w \in \mathbb{R} \right\}.$$

Is S is a subspace of $\mathbb{R}^4$? If so, explain why, give a basis for the subspace and find its dimension.

Exercise $4.8.E.24$

If you have 5 vectors in $\mathbb{R}^5$ and the vectors are linearly independent, can it always be concluded they span $\mathbb{R}^5$? Explain.

Answer available at the end of this section.

Exercise $4.8.E.25$

If you have 6 vectors in $\mathbb{R}^5$, is it possible they are linearly independent? Explain.

Answer available at the end of this section.

Exercise $4.8.E.26$

Suppose A is an $m \times n$ matrix and $\{\vec{w}_1, \cdots, \vec{w}_k\}$ is a linearly independent set of vectors in $A(\mathbb{R}^n) \subseteq \mathbb{R}^m$. Now suppose $A\vec{z}_i = \vec{w}_i$. Show $\{\vec{z}_1, \cdots, \vec{z}_k\}$ is also independent.

Answer available at the end of this section.

Exercise $4.8.E.27$

Let V and W be subspaces of $\mathbb{R}^n$. Show that the intersection $V \cap W$ is also a subspace of $\mathbb{R}^n$

Answers

Exercise $4.8.E.9$

No. Let $\vec{u} = \begin{bmatrix} \frac{\pi}{2} \\ 0 \\ 0 \\ 0 \end{bmatrix}$. Then $2\vec{u} \not\in M$ although $\vec{u} \in M$.

Exercise $4.8.E.10$

No. $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \in M$ but $10 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \not\in M$.

Exercise $4.8.E.11$

This is not a subspace. $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ is in it. However, $(-1)\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ is not.

Exercise $4.8.E.12$

This is a subspace because it is closed with respect to vector addition and scalar multiplication.

Exercise $4.8.E.13$

Yes, this is a subspace because it is closed with respect to vector addition and scalar multiplication.

Exercise $4.8.E.14$

This is not a subspace. $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ is in it. However $(-1)\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}$ is not.

Exercise $4.8.E.15$

This is a subspace. It is closed with respect to vector addition and scalar multiplication.

Exercise $4.8.E.24$

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 $4.8.E.25$

They can't be.

Exercise $4.8.E.26$

Say $\sum_{i=1}^{k} c_i \vec{z}_i = \vec{0}$. Then apply A to it as follows.

$$\sum_{i=1}^{k} c_a A\vec{z}_i = \sum_{i=1}^{k} c_i \vec{w}_i = \vec{0}$$

and so, by linear independence of the $\vec{w}_i$, it follows that each $c_i = 0$.

This page titled 4.8.E: Exercise for Section 4.6 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

4.9: Row, Column and Null Spaces

Learning Objectives

- Determine the row space, column space, and null space of a matrix.
- Calculate the rank and nullity of a matrix and understand their relationship through the Rank-Nullity Theorem.
- Verify if a set of vectors forms a basis for the row or column space.

We begin this section with a new definition.

Definition 4.9.1: Row and Column Space

Let A be an $m \times n$ matrix. The **column space** of A, written $\mathrm{col}(A)$, is the span of the column vectors of A. The **row space** of A, written $\mathrm{row}(A)$, is the span of the row vectors of A.

Using the reduced row-echelon form, we can obtain an efficient description of the row and column space of a matrix. Consider the following lemma.

Lemma 4.9.1: 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 $[\mathrm{column}]$ operations. Then $\mathrm{row}(A) = \mathrm{row}(B)$ $[\mathrm{col}(A) = \mathrm{col}(B)]$.

Proof available at the end of this section.

Consider the following lemma.

Lemma 4.9.2: 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 $\mathrm{row}(R)$, and consequently of $\mathrm{row}(A)$.

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 $\mathrm{col}(A)$.

Before proceeding to an example of this concept, we revisit the definition of rank.

Definition 4.9.2: Rank of a Matrix

Previously, we defined $\mathrm{rank}(A)$ 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:

$$\mathrm{rank}(A) = \dim(\mathrm{row}(A))$$

Consider the following example.

Example 4.9.1: Rank, Column and Row Space

Find the rank of the following matrix and describe the column and row spaces.

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

Solution

The reduced row-echelon form of A is

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

Therefore, the rank is 2.

Notice that the first two columns of R are pivot columns. By the discussion following Lemma 4.9.2, we find the corresponding columns of A, in this case the first two columns. Therefore a basis for $\mathrm{col}(A)$ is given by

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 7 \end{bmatrix} \right\}$$

For example, consider the third column of the original matrix. It can be written as a linear combination of the first two columns of the original matrix as follows.

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

What about an efficient description of the row space? By Lemma 4.9.2 we know that the nonzero rows of R create a basis of $\mathrm{row}(A)$. For the above matrix, the row space equals

$$\mathrm{row}(A) = \mathrm{span}\left\{ \begin{bmatrix} 1 & 0 & -9 & 9 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 5 & -3 & 0 \end{bmatrix} \right\}$$

Notice that the column space of A is given as the span of columns of the original matrix, while the row space of A is the span of rows of the reduced row-echelon form of A.

Consider another example.

Example 4.9.2: 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 $4.9.2$, 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 $4.9.1$: 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 $4.9.1$: 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 $4.9.3$: 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 4.9.3: 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 4.9.4: Image of A

The **image** of A, written $\text{im}\,(A)$ is given by

$$\text{im}\,(A) = \left\{A\vec{x} : \vec{x} \in \mathbb{R}^n\right\}$$

Consider A as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ whose action is given by multiplication. The following diagram displays this scenario.

$$\begin{matrix} \text{null}(A) & & \text{im}(A) \\ \mathbb{R}^n & \overset{A}{\to} & \mathbb{R}^m \end{matrix}$$

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 4.9.4: 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 4.9.2: 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 4.9.5: 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

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

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

$$\begin{bmatrix} -3t \\ t \\ t \end{bmatrix} : t^{\text{int}}\mathbb{R}$$

which can be written as

$$t \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} : t^{\text{int}}\mathbb{R}$$

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 $\left\{A\vec{x} : \vec{x}^{\text{int}}\mathbb{R}^n\right\}$ 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 4.9.6: 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

$$\begin{bmatrix} 1 & 2 & 1 & 0 & 1 & 0 \\ 2 & -1 & 1 & 3 & 0 & 0 \\ 3 & 1 & 2 & 3 & 1 & 0 \\ 4 & -2 & 2 & 6 & 0 & 0 \end{bmatrix} \to \cdots \to \begin{bmatrix} 1 & 0 & \frac{3}{5} & \frac{6}{5} & \frac{1}{5} & 0 \\ 0 & 1 & \frac{1}{5} & -\frac{3}{5} & \frac{2}{5} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

It follows that the first two columns are pivot columns, and the next three correspond to parameters. Therefore, $\text{null}\,(A)$ is given by

$$\begin{bmatrix} \left(-\frac{3}{5}\right)s + \left(-\frac{6}{5}\right)t + \left(\frac{1}{5}\right)r \\ \left(-\frac{1}{5}\right)s + \left(\frac{3}{5}\right)t + \left(-\frac{2}{5}\right)r \\ s \\ t \\ r \end{bmatrix} : s, t, r^{\text{int}}\mathbb{R}.$$

We write this in the form

$$s\begin{bmatrix} -\frac{3}{5} \\ -\frac{1}{5} \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} -\frac{6}{5} \\ \frac{3}{5} \\ 0 \\ 1 \\ 0 \end{bmatrix} + r\begin{bmatrix} \frac{1}{5} \\ -\frac{2}{5} \\ 0 \\ 0 \\ 1 \end{bmatrix} : s, t, r^{\text{int}}\mathbb{R}.$$

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 4.9.5: 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, sometimes known as the **Rank-Nullity Theorem**.

Theorem 4.9.3: Rank and Nullity

Let A be an $m \times n$ matrix. Then $\text{rank}\,(A) + \dim(\text{null}\,(A)) = n$.

Example 4.9.7: 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 4.9.5 we determined that the reduced row-echelon form of A is given by

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

Therefore the rank of A is 2. We also determined that the null space of A is given by

$$\text{null}(A) = \text{span}\left\{ \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \right\}$$

Therefore the nullity of A is 1. It follows from Theorem 4.9.3 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 4.9.4

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 $\vec{x}^{\text{int}}\mathbb{R}^n$, then $\vec{x} = \vec{0}_n$.

Theorem 4.9.5

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 $\vec{b}^{\text{int}}\mathbb{R}^m$.

Answers

Lemma $4.9.1$**: Effect of Row Operations on Row Space**

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

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, $\text{row}(B) = \text{row}(A)$.

This page titled 4.9: Row, Column and Null Spaces is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

- 4.10: Spanning, Linear Independence and Basis in Rⁿ by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.9.E: Exercise for Section 4.7

Exercise 4.9.E.1

Given the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 0 & 1 & -1 & 2 \end{bmatrix}$$

a. Find a basis for the row space of A.
b. Find a basis for the column space of A.
c. Find a basis for the null space of A.

Answer available at the end of this section.

Exercise 4.9.E.2

Let

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

a. Compute the rank of B.
b. Compute the nullity of B.
c. Verify the Rank-Nullity Theorem for this matrix.

Answer available at the end of this section.

Exercise 4.9.E.3

Given the matrix

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

a. Compute the rank of D.
b. Compute the nullity of D.
c. Find a basis for the null space of D.
d. Does the Rank-Nullity Theorem hold for D?

Answer available at the end of this section.

Exercise 4.9.E.4

Consider the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 9 & 12 \end{bmatrix}. \tag{4.9.E.1}$$

a. Find a basis for the row space of A.
b. Find a basis for the column space of A.
c. Find a basis for the null space of A.
d. Compute rank and nullity of A, and verify the Rank-Nullity Theorem.

Exercise $4.9.E.5$

Let

$$B = \begin{bmatrix} 1 & -1 & 3 & 2 & 4 \\ 2 & 0 & 5 & 6 & 8 \\ 3 & -1 & 7 & 8 & 12 \end{bmatrix}$$

a. Compute the rank of B.
b. Compute the nullity of B.

Answer available at the end of this section.

Exercise $4.9.E.6$

Consider the matrix

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

Find a basis for the null space of D

Exercise $4.9.E.7$

Show that if A is an $m \times n$ matrix, then $\ker(A)$ is a subspace of $\mathbb{R}^n$.

Answer available at the end of this section.

Exercise $4.9.E.8$

Find the rank of the following matrix. Also find a basis for the row and column spaces.

$$\begin{bmatrix} 1 & 3 & 0 & -2 & 0 & 3 \\ 3 & 9 & 1 & -7 & 0 & 8 \\ 1 & 3 & 1 & -3 & 1 & -1 \\ 1 & 3 & -1 & -1 & -2 & 10 \end{bmatrix}$$

Answer available at the end of this section.

Exercise $4.9.E.9$

Find the rank of the following matrix. Also find a basis for the row and column spaces.

$$\begin{bmatrix} 1 & 3 & 0 & -2 & 7 & 3 \\ 3 & 9 & 1 & -7 & 23 & 8 \\ 1 & 3 & 1 & -3 & 9 & 2 \\ 1 & 3 & -1 & -1 & 5 & 4 \end{bmatrix}$$

Exercise 4.9.E.10

Find the rank of the following matrix. Also find a basis for the row and column spaces.

$$\begin{bmatrix} 1 & 0 & 3 & 0 & 7 & 0 \\ 3 & 1 & 10 & 0 & 23 & 0 \\ 1 & 1 & 4 & 1 & 7 & 0 \\ 1 & -1 & 2 & -2 & 9 & 1 \end{bmatrix}$$

Exercise 4.9.E.11

Find the rank of the following matrix. Also find a basis for the row and column spaces.

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

Exercise 4.9.E.12

Find the rank of the following matrix. Also find a basis for the row and column spaces.

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

Exercise 4.9.E.13

Find the rank of the following matrix. Also find a basis for the row and column spaces.

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

Exercise 4.9.E.14

Find $\ker(A)$ and $\operatorname{im}(A)$ for the following matrices.

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

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

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

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

Exercise $4.9.E.15$

Suppose A is an $m \times n$ matrix and B is an $n \times p$ matrix. Show that

$$\dim(\ker(AB)) \leq \dim(\ker(A)) + \dim(\ker(B)).$$

Consider the subspace, $B(\mathbb{R}^p) \cap \ker(A)$ and suppose a basis for this subspace is $\{\vec{w}_1, \cdots, \vec{w}_k\}$. Now suppose $\{\vec{u}_1, \cdots, \vec{u}_r\}$ is a basis for $\ker(B)$. Let $\{\vec{z}_1, \cdots, \vec{z}_k\}$ be such that $B\vec{z}_1 = \vec{w}_i$ and argue that

$$\ker(AB) \subseteq span\{\vec{u}_1, \cdots, \vec{u}_r, \vec{z}_1, \cdots, \vec{z}_k\}.$$

Answer available at the end of this section.

Exercise $4.9.E.16$

Suppose A is an m×nm \times nm×n matrix. Prove that every vector in $\mathbb{R}^n$ can be uniquely written as a sum of a vector in the row space of A and a vector in the null space of A.

Exercise $4.9.E.17$

Let A and B be two $m \times n$ matrices such that BBB is obtained from A by elementary row operations. Prove that A and B have the same row space.

Answers

Exercise $4.9.E.1$

Given the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 0 & 1 & -1 & 2 \end{bmatrix}$$

a. Basis for the row space: The row space is spanned by the linearly independent rows of A after row reduction:

$$\text{Row space basis: } \{[1 \quad 2 \quad 3 \quad 4], [0 \quad 1 \quad -1 \quad 2]\}.$$

b. Basis for the column space: The column space is spanned by the pivot columns of A:

$$\text{Column space basis: } \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} \right\}.$$

c. Basis for the null space: Solve $A\mathbf{x} = 0$ to find the null space basis:

$$\text{Null space basis: } \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Exercise $4.9.E.2$

Let

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

a. Rank of B: The rank is the number of linearly independent rows:

$$\text{Rank}(B) = 1.$$

b. Nullity of B: The nullity is computed as:

$$\text{Nullity}(B) = 3 - 1 = 2.$$

c. Verification of Rank-Nullity Theorem:

$$\text{Rank}(B) + \text{Nullity}(B) = 1 + 2 = 3,$$

which matches the number of columns, so the theorem holds.

Exercise $4.9.E.3$

Given the matrix

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

\textbf{Rank of D:} \ The rank is the number of linearly independent rows:

$$\text{Rank}(D) = 2.$$

\textbf{Nullity of D:} \

$$\text{Nullity}(D) = 4 - 2 = 2.$$

\textbf{Basis for the null space:} \ Solve $D\mathbf{x} = 0$:

$$\text{Null space basis: } \left\{ \begin{bmatrix} 1 \\ -5 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -4 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

\textbf{Verification of Rank-Nullity Theorem:} \

$$\text{Rank}(D) + \text{Nullity}(D) = 2 + 2 = 4,$$

which matches the number of columns, so the theorem holds.

Exercise $4.9.E.5$

a. Row reduce B:

$$\begin{bmatrix} 1 & -1 & 3 & 2 & 4 \\ 2 & 0 & 5 & 6 & 8 \\ 3 & -1 & 7 & 8 & 12 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -1 & 3 & 2 & 4 \\ 0 & 2 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{Rank}(B) = 2$$

b. Using the Rank-Nullity Theorem:

$$\text{Nullity}(B) = 5 - 2 = 3$$

Exercise $4.9.E.7$

If $\vec{x}$, $\vec{y}^{\text{int}}\ker(A)$, then $A(\vec{x}) = \vec{0}$ and $A(\vec{y}) = \vec{0}$. Now let's look at $a\vec{x} + b\vec{y}$ for some $a, b^{\text{int}}\mathbb{R}$. Then,

$$A(a\vec{x} + b\vec{y}) = aA\vec{x} + bA\vec{y} = a\vec{0} + b\vec{0} = \vec{0}$$

and so $\ker(A)$ is closed under linear combinations. Hence it is a subspace.

Exercise $4.9.E.8$

The rank of the given matrix is 3. A basis for the row space is:

$$\{[1 \quad 3 \quad 0 \quad -2 \quad 0 \quad 3], [0 \quad 0 \quad 1 \quad -1 \quad 1 \quad -3], [0 \quad 0 \quad 0 \quad 0 \quad 1 \quad -5]\}$$

A basis for the column space is:

$$\left\{ \begin{bmatrix} 1 \\ 3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 9 \\ 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ -1 \end{bmatrix} \right\}$$

Exercise $4.9.E.15$

Here is how you do this. Suppose $AB\vec{x} = \vec{0}$. Then $B\vec{x} \in \ker(A) \cap B(\mathbb{R}^p)$ and so $B\vec{x} = \sum_{i=1}^{k} B\vec{z}_i$ showing that

$$\vec{x} - \sum_{i=1}^{k} \vec{z}_i^{\,\text{int}}\ker(B)$$

Consider $B(\mathbb{R}^p) \cap \ker(A)$ and let a basis be $\{\vec{w}_1, \cdots, \vec{w}_k\}$. Then each $\vec{w}_i$ is of the form $B\vec{z}_i = \vec{w}_i$. Therefore, $\{\vec{z}_1, \cdots, \vec{z}_k\}$ is linearly independent and $AB\vec{z}_i = 0$. Now let $\{\vec{u}_1, \cdots, \vec{u}_r\}$ be a basis for $\ker(B)$. If $AB\vec{x} = \vec{0}$, then $B\vec{x} \in \ker(A) \cap B(\mathbb{R}^p)$ and so $B\vec{x} = \sum_{i=1}^{k} c_i B\vec{z}_1$ which implies

$$\vec{x} - \sum_{i=1}^{k} c_i \vec{z}_i^{\,\text{int}}\ker(B)$$

and so it is of the form

$$\vec{x} - \sum_{i=1}^{k} c_i \vec{z}_i = \sum_{j=1}^{r} d_j \vec{u}_j$$

It follows that if $AB\vec{x} = \vec{0}$ so that $\vec{x} \in \ker(AB)$, then

$$\vec{x}^{\,\text{int}} span(\vec{z}_1, \cdots, \vec{z}_k, \vec{u}_1, \cdots, \vec{u}_r).$$

Therefore,

$$\begin{aligned} \dim(\ker(AB)) &\leq k + r = \dim(B(\mathbb{R}^p) \cap \ker(A)) + \dim(\ker(B)) \\ &\leq \dim(\ker(A)) + \dim(\ker(B)) \end{aligned} \tag{4.9.E.2}$$

This page titled 4.9.E: Exercise for Section 4.7 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

4.10: 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:

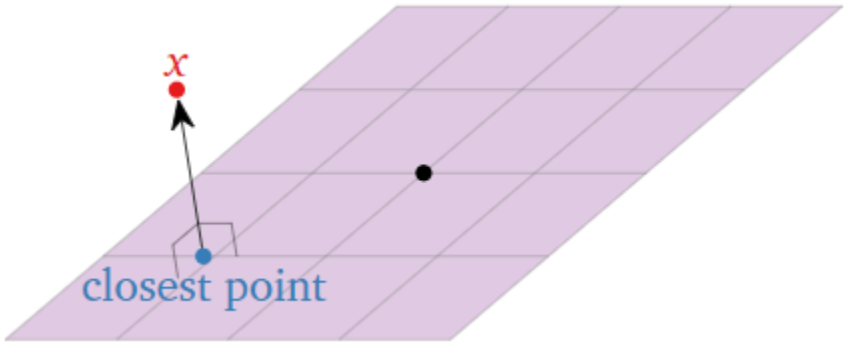


Figure 4.10.1

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 4.10.1: 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^Ty.

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 4.10.1: 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:

- $x \cdot x \geq 0$
- $x \cdot x = 0 \leftrightarrow x = 0$.

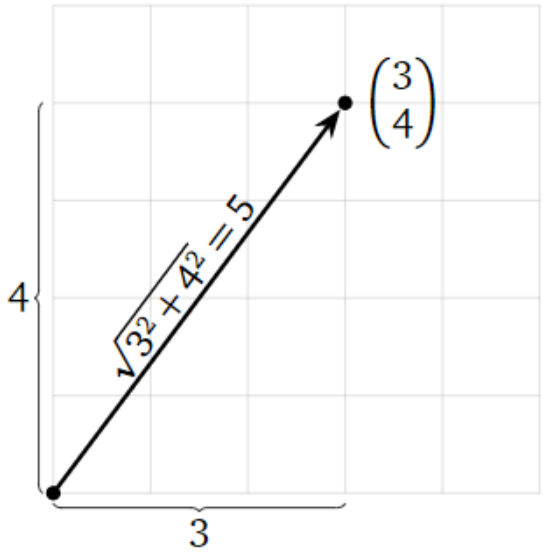
This leads to a good definition of *length*.

Fact 4.10.1

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.



$$\left\| \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right\| = \sqrt{3^2 + 4^2} = 5$$

Figure 4.10.1

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 4.10.1

Suppose that $\|x\| = 2, \quad \|y\| = 3$, 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 4.10.2

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 4.10.2: 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:

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

Example 4.10.2

Find the distance from $(1, 2)$ to $(4, 4)$.

Solution

Let $x = (1, 2)$ and $y = (4, 4)$. Then

$$\operatorname{dist}(x, y) = \|y - x\| = \left\| \begin{pmatrix} 3 \\ 2 \end{pmatrix} \right\| = \sqrt{3^2 + 2^2} = \sqrt{13}.$$

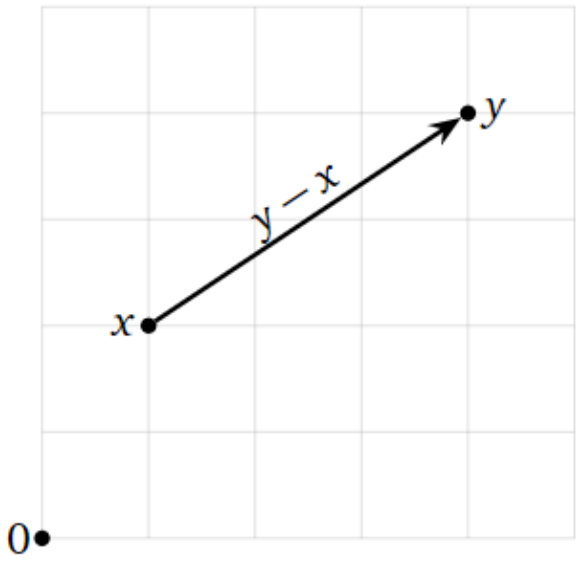


Figure 4.10.2

Vectors with length 1 are very common in applications, so we give them a name.

Definition 4.10.3: 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 4.10.3

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 4.10.3

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:

$$u = \frac{x}{\|x\|} = \frac{1}{\sqrt{3^2+4^2}} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 3/5 \\ 4/5 \end{pmatrix}.$$

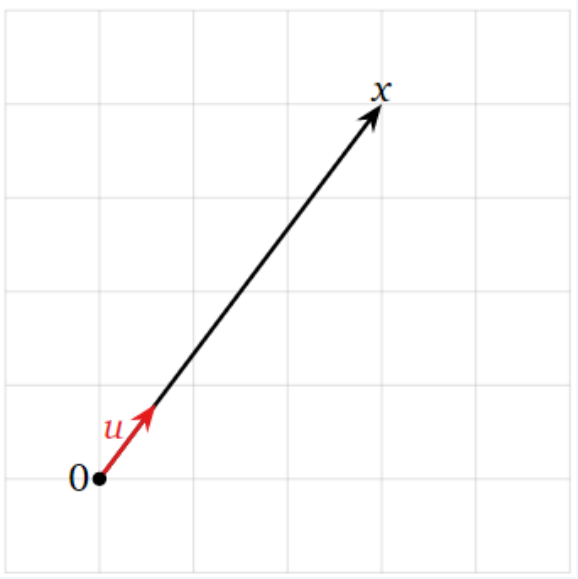


Figure 4.10.3

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 4.10.4: 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 4.10.2

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

$$\|y - x\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\|\,\|y\| \cos\alpha.$$

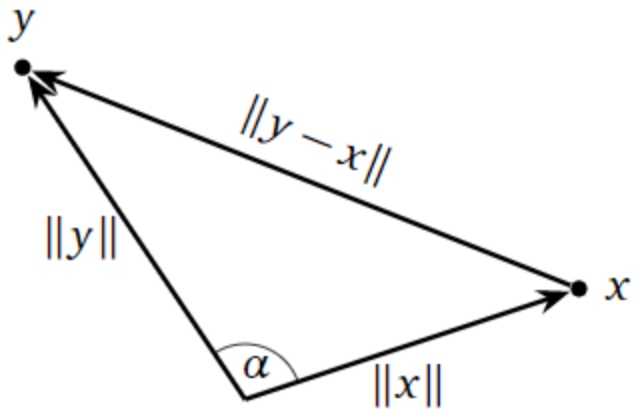


Figure 4.10.4

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,

$$\begin{aligned} x \text{ and } y \text{ are perpendicular} &\leftrightarrow \|x\|^2 + \|y\|^2 = \|y-x\|^2 \\ &\leftrightarrow x\cdot x + y\cdot y = (y-x)\cdot(y-x) \\ &\leftrightarrow x\cdot x + y\cdot y = y\cdot y + x\cdot x - 2x\cdot y \\ &\leftrightarrow x\cdot y = 0. \end{aligned}$$

To reiterate:

Note 4.10.3

$$x \perp y \quad \leftrightarrow \quad x\cdot y = 0 \quad \leftrightarrow \quad \|y-x\|^2 = \|x\|^2 + \|y\|^2.$$

Example 4.10.4

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 4.10.5

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

$$\begin{aligned} 0 = x \cdot v &= \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = x_1 + x_2 - x_3 \\ 0 = x \cdot w &= \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = x_1 + x_2 + x_3. \end{aligned} \tag{4.10.1}$$

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 4.10: Dot Products and Orthogonality is shared under a GNU FDL license and was authored, remixed, and/or curated by Dan Margalit and 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 and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

4.11: Orthogonal Vectors and Matrices

Learning Objectives

- Determine if a given set is orthogonal or orthonormal.
- Determine if a given matrix is orthogonal.

In this section, we examine what it means for vectors (and sets of vectors) to be orthogonal and orthonormal.

Recall from the properties of the dot product of vectors that two vectors $\vec{u}$ and $\vec{v}$ are **orthogonal** if $\vec{u} \cdot \vec{v} = 0$. Suppose a vector is orthogonal to a spanning set of $\mathbb{R}^n$. What can be said about such a vector? This is the discussion in the following example.

Example 4.11.1: Orthogonal Vector to a Spanning Set

Let $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}^{\text{int}}\mathbb{R}^n$ and suppose $\mathbb{R}^n = \text{span}\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$. Furthermore, suppose that there exists a vector $\vec{u}^{\text{int}}\mathbb{R}^n$ for which $\vec{u} \cdot \vec{x}_j = 0$ for all j, $1 \leq j \leq k$. What type of vector is $\vec{u}$?

Solution

Write $\vec{u} = t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k$ for some $t_1, t_2, \ldots, t_k^{\text{int}}\mathbb{R}$ (this is possible because $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$ span $\mathbb{R}^n$).

Then

$$\begin{aligned} \|\vec{u}\|^2 &= \vec{u} \cdot \vec{u} \\ &= \vec{u} \cdot (t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k) \\ &= \vec{u} \cdot (t_1\vec{x}_1) + \vec{u} \cdot (t_2\vec{x}_2) + \cdots + \vec{u} \cdot (t_k\vec{x}_k) \\ &= t_1(\vec{u} \cdot \vec{x}_1) + t_2(\vec{u} \cdot \vec{x}_2) + \cdots + t_k(\vec{u} \cdot \vec{x}_k) \\ &= t_1(0) + t_2(0) + \cdots + t_k(0) = 0. \end{aligned} \tag{4.11.1}$$

Since $\|\vec{u}\|^2 = 0$, $\|\vec{u}\| = 0$. We know that $\|\vec{u}\| = 0$ if and only if $\vec{u} = \vec{0}_n$. Therefore, $\vec{u} = \vec{0}_n$. In conclusion, the only vector orthogonal to every vector of a spanning set of $\mathbb{R}^n$ is the zero vector.

We can now discuss what is meant by an orthogonal set of vectors.

Definition 4.11.1: Orthogonal Set of Vectors

Let $\{\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_m\}$ be a set of vectors in $\mathbb{R}^n$. Then this set is called an **orthogonal set** if the following conditions hold:

1. $\vec{u}_i \cdot \vec{u}_j = 0$ for all $i \neq j$
2. $\vec{u}_i \neq \vec{0}$ for all i

If we have an orthogonal set of vectors and normalize each vector so they have length 1, the resulting set is called an **orthonormal set** of vectors. They can be described as follows.

Definition 4.11.2: Orthonormal Set of Vectors

A set of vectors, $\{\vec{w}_1, \cdots, \vec{w}_m\}$ is said to be an **orthonormal** set if

$$\vec{w}_i \cdot \vec{w}_j = \delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

Note that all orthonormal sets are orthogonal, but the reverse is not necessarily true since the vectors may not be normalized. In order to normalize the vectors, we simply need divide each one by its length.

Definition 4.11.3: Normalizing an Orthogonal Set

Normalizing an orthogonal set is the process of turning an orthogonal (but not orthonormal) set into an orthonormal set. If $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is an orthogonal subset of $\mathbb{R}^n$, then

$$\left\{ \frac{1}{\|\vec{u}_1\|}\vec{u}_1, \frac{1}{\|\vec{u}_2\|}\vec{u}_2, \ldots, \frac{1}{\|\vec{u}_k\|}\vec{u}_k \right\}$$

is an orthonormal set.

We illustrate this concept in the following example.

Example 4.11.2: Orthonormal Set

Consider the set of vectors given by

$$\{\vec{u}_1, \vec{u}_2\} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$$

Show that it is an orthogonal set of vectors but not an orthonormal one. Find the corresponding orthonormal set.

Solution

One easily verifies that $\vec{u}_1 \cdot \vec{u}_2 = 0$ and $\{\vec{u}_1, \vec{u}_2\}$ is an orthogonal set of vectors. On the other hand one can compute that $\|\vec{u}_1\| = \|\vec{u}_2\| = \sqrt{2} \neq 1$ and thus it is not an orthonormal set.

Thus to find a corresponding orthonormal set, we simply need to normalize each vector. We will write $\{\vec{w}_1, \vec{w}_2\}$ for the corresponding orthonormal set. Then,

$$\begin{aligned} \vec{w}_1 &= \frac{1}{\|\vec{u}_1\|}\vec{u}_1 \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \end{aligned} \tag{4.11.2}$$

Similarly,

$$\begin{aligned} \vec{w}_2 &= \frac{1}{\|\vec{u}_2\|}\vec{u}_2 \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \end{aligned} \tag{4.11.3}$$

Therefore the corresponding orthonormal set is

$$\{\vec{w}_1, \vec{w}_2\} = \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right\}$$

You can verify that this set is orthogonal.

Consider an orthogonal set of vectors in $\mathbb{R}^n$, written $\{\vec{w}_1, \cdots, \vec{w}_k\}$ with $k \leq n$. The span of these vectors is a subspace W of $\mathbb{R}^n$. If

we could show that this orthogonal set is also linearly independent, we would have a basis of W. We will show this in the next theorem.

Theorem 4.11.1: Orthogonal Basis of a Subspace

Let $\{\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_k\}$ be an orthonormal set of vectors in $\mathbb{R}^n$. Then this set is linearly independent and forms a basis for the subspace $W = \text{span}\{\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_k\}$.

Proof available at the end of this section.

If an orthogonal set is a basis for a subspace, we call this an orthogonal basis. Similarly, if an orthonormal set is a basis, we call this an orthonormal basis.

We conclude this section with a discussion of Fourier expansions. Given any orthogonal basis B of $\mathbb{R}^n$ and an arbitrary vector $\vec{x}^{\text{int}}\mathbb{R}^n$, how do we express $\vec{x}$ as a linear combination of vectors in B? The solution is Fourier expansion.

Theorem 4.11.2: Fourier Expansion

Let V be a subspace of $\mathbb{R}^n$ and suppose $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}$ is an orthogonal basis of V. Then for any $\vec{x}^{\text{int}}V$,

$$\vec{x} = \left(\frac{\vec{x}\cdot\vec{u}_1}{\|\vec{u}_1\|^2}\right)\vec{u}_1 + \left(\frac{\vec{x}\cdot\vec{u}_2}{\|\vec{u}_2\|^2}\right)\vec{u}_2 + \cdots + \left(\frac{\vec{x}\cdot\vec{u}_m}{\|\vec{u}_m\|^2}\right)\vec{u}_m$$

This expression is called the **Fourier expansion** of $\vec{x}$, and

$$\frac{\vec{x}\cdot\vec{u}_j}{\|\vec{u}_j\|^2},$$

$j = 1, 2, \ldots, m$ are the Fourier coefficients.

Consider the following example.

Example 4.11.3: Fourier Expansion

Let $\vec{u}_1 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$, $\vec{u}_2 = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$, and $\vec{u}_3 = \begin{bmatrix} 5 \\ 1 \\ -2 \end{bmatrix}$, and let $\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Then $B = \{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ is an orthogonal basis of $\mathbb{R}^3$.

Compute the Fourier expansion of $\vec{x}$, thus writing $\vec{x}$ as a linear combination of the vectors of B.

Solution

Since B is a basis (verify!) there is a unique way to express $\vec{x}$ as a linear combination of the vectors of B. Moreover since B is an orthogonal basis (verify!), then this can be done by computing the Fourier expansion of $\vec{x}$.

That is:

$$\vec{x} = \left(\frac{\vec{x}\cdot\vec{u}_1}{\|\vec{u}_1\|^2}\right)\vec{u}_1 + \left(\frac{\vec{x}\cdot\vec{u}_2}{\|\vec{u}_2\|^2}\right)\vec{u}_2 + \left(\frac{\vec{x}\cdot\vec{u}_3}{\|\vec{u}_3\|^2}\right)\vec{u}_3.$$

We readily compute:

$$\frac{\vec{x}\cdot\vec{u}_1}{\|\vec{u}_1\|^2} = \frac{2}{6}, \quad \frac{\vec{x}\cdot\vec{u}_2}{\|\vec{u}_2\|^2} = \frac{3}{5}, \quad \text{and} \quad \frac{\vec{x}\cdot\vec{u}_3}{\|\vec{u}_3\|^2} = \frac{4}{30}.$$

Therefore,

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + \frac{3}{5}\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} + \frac{2}{15}\begin{bmatrix} 5 \\ 1 \\ -2 \end{bmatrix}.$$

Recall that the process to find the inverse of a matrix was often cumbersome. In contrast, it was very easy to take the transpose of a matrix. Luckily for some special matrices, the transpose equals the inverse. When an $n \times n$ matrix has all real entries and its transpose equals its inverse, the matrix is called an **orthogonal matrix**.

The precise definition is as follows.

Definition 4.11.4: Orthogonal Matrices

A real $n \times n$ matrix U is called an **orthogonal** matrix if

$$UU^T = U^T U = I.$$

Note since U is assumed to be a square matrix, it suffices to verify only one of these equalities $UU^T = I$ or $U^T U = I$ holds to guarantee that U^T is the inverse of U.

Consider the following example.

Example 4.11.4: Orthogonal Matrix

Orthogonal Matrix Show the matrix

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

is orthogonal.

Solution

All we need to do is verify (one of the equations from) the requirements of Definition 4.11.4.

$$UU^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Since $UU^T = I$, this matrix is orthogonal.

Here is another example.

Example 4.11.5: Orthogonal Matrix

Orthogonal Matrix Let $U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}$. Is U orthogonal?

Solution

Again the answer is yes and this can be verified simply by showing that $U^T U = I$:

$$\begin{aligned} U^TU &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \tag{4.11.4}$$

When we say that U is orthogonal, we are saying that $UU^T = I$, meaning that

$$\sum_j u_{ij}u_{jk}^T = \sum_j u_{ij}u_{kj} = \delta_{ik}$$

where δ_{ij} is the **Kronecker delta** defined by

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

In words, the product of the ith row of U with the kth row gives 1 if $i = k$ and 0 if $i \neq k$. The same is true of the columns because $U^TU = I$ also. Therefore,

$$\sum_j u_{ij}^T u_{jk} = \sum_j u_{ji}u_{jk} = \delta_{ik}$$

which says that the product of one column with another column gives 1 if the two columns are the same and 0 if the two columns are different.

More succinctly, this states that if $\vec{u}_1, \cdots, \vec{u}_n$ are the columns of U, an orthogonal matrix, then

$$\vec{u}_i \cdot \vec{u}_j = \delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

We will say that the columns form an orthonormal set of vectors, and similarly for the rows. Thus a matrix is **orthogonal** if its rows (or columns) form an **orthonormal** set of vectors. Notice that the convention is to call such a matrix orthogonal rather than orthonormal (although this may make more sense!).

Proposition 4.11.1: Orthonormal Basis

The rows of an $n \times n$ orthogonal matrix form an orthonormal basis of $\mathbb{R}^n$. Further, any orthonormal basis of $\mathbb{R}^n$ can be used to construct an $n \times n$ orthogonal matrix.

Proof available at the end of this section.

Consider the following proposition.

Proposition 4.11.2: Determinant of Orthogonal Matrices

Suppose U is an orthogonal matrix. Then $\det(U) = \pm 1$.

Proof available at the end of this section.

Orthogonal matrices are divided into two classes, proper and improper. The proper orthogonal matrices are those whose determinant equals 1 and the improper ones are those whose determinant equals -1. The reason for the distinction is that the improper orthogonal matrices are sometimes considered to have no physical significance. These matrices cause a change in orientation which would correspond to material passing through itself in a non physical manner. Thus in considering which

coordinate systems must be considered in certain applications, you only need to consider those which are related by a proper orthogonal transformation. Geometrically, the linear transformations determined by the proper orthogonal matrices correspond to the composition of rotations.

We conclude this section with two useful properties of orthogonal matrices.

Example 4.11.6: Product and Inverse of Orthogonal Matrices

Suppose A and B are orthogonal matrices. Then AB and A^{-1} both exist and are orthogonal.

Solution

First we examine the product AB.

$$(AB)(B^TA^T) = A(BB^T)A^T = AA^T = I$$

Since AB is square, $B^TA^T = (AB)^T$ is the inverse of AB, so AB is invertible, and $(AB)^{-1} = (AB)^T$ Therefore, AB is orthogonal.

Next we show that $A^{-1} = A^T$ is also orthogonal.

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

Therefore A^{-1} is also orthogonal.

Answers

Theorem 4.11.1: Orthogonal Basis of a Subspace

To show it is a linearly independent set, suppose a linear combination of these vectors equals $\vec{0}$, such as:

$$a_1\vec{w}_1 + a_2\vec{w}_2 + \cdots + a_k\vec{w}_k = \vec{0}, a_i^{\text{int}}\mathbb{R}$$

We need to show that all $a_i = 0$. To do so, take the dot product of each side of the above equation with the vector $\vec{w}_i$ and obtain the following.

$$\begin{aligned} \vec{w}_i \cdot (a_1\vec{w}_1 + a_2\vec{w}_2 + \cdots + a_k\vec{w}_k) &= \vec{w}_i \cdot \vec{0} \\ a_1(\vec{w}_i \cdot \vec{w}_1) + a_2(\vec{w}_i \cdot \vec{w}_2) + \cdots + a_k(\vec{w}_i \cdot \vec{w}_k) &= 0 \end{aligned} \tag{4.11.5}$$

Now since the set is orthogonal, $\vec{w}_i \cdot \vec{w}_m = 0$ for all $m \neq i$, so we have:

$$a_1(0) + \cdots + a_i(\vec{w}_i \cdot \vec{w}_i) + \cdots + a_k(0) = 0$$

$$a_i\|\vec{w}_i\|^2 = 0$$

Since the set is orthogonal, we know that $\|\vec{w}_i\|^2 \neq 0$. It follows that $a_i = 0$. Since the a_i was chosen arbitrarily, the set $\{\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_k\}$ is linearly independent.

Finally since $W = \text{span}\{\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_k\}$, the set of vectors also spans W and therefore forms a basis of W.

Proposition 4.11.1: Orthonormal Basis

Recall from Theorem 4.11.1 that an orthonormal set is linearly independent and forms a basis for its span. Since the rows of an $n \times n$ orthogonal matrix form an orthonormal set, they must be linearly independent. Now we have n linearly independent vectors, and it follows that their span equals $\mathbb{R}^n$. Therefore these vectors form an orthonormal basis for $\mathbb{R}^n$.

Suppose now that we have an orthonormal basis for $\mathbb{R}^n$. Since the basis will contain n vectors, these can be used to construct an $n \times n$ matrix, with each vector becoming a row. Therefore the matrix is composed of orthonormal rows, which by our above discussion, means that the matrix is orthogonal. Note we could also have construct a matrix with each vector becoming a column instead, and this would again be an orthogonal matrix. In fact this is simply the transpose of the previous matrix.

Proposition 4.11.2: Determinant of Orthogonal Matrices

This result follows from the properties of determinants. Recall that for any matrix A, $\det(A)^T = \det(A)$. Now if U is orthogonal, then:

$$(\det(U))^2 = \det(U^T)\det(U) = \det(U^TU) = \det(I) = 1$$

Therefore $(\det(U))^2 = 1$ and it follows that $\det(U) = \pm 1$.

This page titled 4.11: Orthogonal Vectors and Matrices is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

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

4.11.E: Exercise for Section 4.8

Exercise 4.11.E.1

Determine whether the following set of vectors is orthogonal. If it is orthogonal, determine whether it is also orthonormal.

$$\begin{bmatrix} \frac{1}{6}\sqrt{2}\sqrt{3} \\ \frac{1}{3}\sqrt{2}\sqrt{3} \\ -\frac{1}{6}\sqrt{2}\sqrt{3} \end{bmatrix}, \begin{bmatrix} \frac{1}{2}\sqrt{2} \\ 0 \\ \frac{1}{2}\sqrt{2} \end{bmatrix}, \begin{bmatrix} -\frac{1}{3}\sqrt{3} \\ \frac{1}{3}\sqrt{3} \\ \frac{1}{3}\sqrt{3} \end{bmatrix}$$

If the set of vectors is orthogonal but not orthonormal, give an orthonormal set of vectors which has the same span.

Exercise 4.11.E.2

Determine whether the following set of vectors is orthogonal. If it is orthogonal, determine whether it is also orthonormal.

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

If the set of vectors is orthogonal but not orthonormal, give an orthonormal set of vectors which has the same span.

Answer available at the end of this section.

Exercise 4.11.E.3

Determine whether the following set of vectors is orthogonal. If it is orthogonal, determine whether it is also orthonormal.

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

If the set of vectors is orthogonal but not orthonormal, give an orthonormal set of vectors which has the same span.

Exercise 4.11.E.4

Determine whether the following set of vectors is orthogonal. If it is orthogonal, determine whether it is also orthonormal.

$$\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

If the set of vectors is orthogonal but not orthonormal, give an orthonormal set of vectors which has the same span.

Exercise 4.11.E.5

Determine whether the following set of vectors is orthogonal. If it is orthogonal, determine whether it is also orthonormal.

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

If the set of vectors is orthogonal but not orthonormal, give an orthonormal set of vectors which has the same span.

Exercise 4.11.E.6

Here are some matrices. Label according to whether they are symmetric, skew symmetric, or orthogonal.

a. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

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

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

Answer available at the end of this section.

Exercise 4.11.E.7

In this exercise, you will show that orthogonal matrices are exactly those which preserve length.

a. For U an orthogonal matrix, explain why $||U\vec{x}|| = ||\vec{x}||$ for any vector $\vec{x}$.
b. Next explain why if U is an $n \times n$ matrix with the property that $||U\vec{x}|| = ||\vec{x}||$ for all vectors, $\vec{x}$, then U must be orthogonal.

Answer available at the end of this section.

Exercise 4.11.E.8

Suppose U is an orthogonal $n \times n$ matrix. Explain why $rank(U) = n$.

Answer available at the end of this section.

Exercise 4.11.E.9

Fill in the missing entries to make the matrix orthogonal.

$$\begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & - & - \\ - & \frac{\sqrt{6}}{3} & - \end{bmatrix}.$$

Answer available at the end of this section.

Exercise 4.11.E.10

Fill in the missing entries to make the matrix orthogonal.

$$\begin{bmatrix} \frac{2}{3} & \frac{\sqrt{2}}{2} & \frac{1}{6}\sqrt{2} \\ \frac{2}{3} & - & - \\ - & 0 & - \end{bmatrix}$$

Answer available at the end of this section.

Exercise 4.11.E.11

Fill in the missing entries to make the matrix orthogonal.

$$\begin{bmatrix} \frac{1}{3} & -\frac{2}{\sqrt{5}} & - \\ \frac{2}{3} & 0 & - \\ - & - & \frac{4}{15}\sqrt{5} \end{bmatrix}$$

Answer available at the end of this section.

Exercise 4.11.E.12

Consider the set of $\mathbb{R}^3$ vectors:

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

a. Show that this set forms an $\mathbf{orthogonal\ basis}$ for $\mathbb{R}^3$.

b. Given the vector $\vec{w} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$, find the Fourier expansion of $\vec{w}$ in terms of this basis.

Exercise 4.11.E.13

Consider the set of $\mathbb{R}^4$ vectors:

$$\vec{u}_1 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{u}_2 = \frac{1}{2}\begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \quad \vec{u}_3 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \quad \vec{u}_4 = \frac{1}{2}\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}.$$

a. Show that this set forms an $\mathbf{orthonormal\ basis}$ for $\mathbb{R}^4$.

b. Given the vector: $\vec{w} = \begin{bmatrix} 3 \\ 1 \\ -2 \\ 0 \end{bmatrix}$, find the Fourier expansion of $\vec{w}$ in terms of this basis.

Exercise 4.11.E.14

Show that if a matrix A has orthogonal columns, then $A^T A$ is a diagonal matrix.

Hint available at the end of this section.

Exercise $4.11.E.15$

Prove that if Q is an orthogonal matrix, then it preserves inner products, i.e., for any vectors $\vec{x}, \vec{y}^{\text{int}}\mathbb{R}^n$,

$$\langle Q\vec{x}, Q\vec{y}\rangle = \langle \vec{x}, \vec{y}\rangle.$$

Hint available at the end of this section.

Answers

Exercise $4.11.E.2$

The given set of vectors:

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

is orthogonal since their pairwise dot products are zero. To normalize them, we compute their norms:

$$\|\vec{v}_1\| = \sqrt{6}, \quad \|\vec{v}_2\| = \sqrt{2}, \quad \|\vec{v}_3\| = \sqrt{3}$$

Thus, an orthonormal basis with the same span is:

$$\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{bmatrix}$$

$$\vec{u}_2 = \frac{\vec{v}_2}{\|\vec{v}_2\|} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\vec{u}_3 = \frac{\vec{v}_3}{\|\vec{v}_3\|} = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

Exercise $4.11.E.6$

a. Orthogonal
b. Symmetric
c. Skew Symmetric

Exercise $4.11.E.7$

a. Suppose U is an orthogonal matrix. $||U\vec{x}||^2 = U\vec{x} \bullet U\vec{x} = U^TU\vec{x} \bullet \vec{x} = I\vec{x} \bullet \vec{x} = ||\vec{x}||^2$.
b. Now, suppose distance is preserved by U. Then

$$\begin{aligned}(U(\vec{x}+\vec{y})) \bullet (U(\vec{x}+\vec{y})) &= ||Ux||^2 + ||Uy||^2 + 2(Ux \bullet Uy) \\ &= ||\vec{x}||^2 + ||\vec{y}||^2 + 2(U^TU\vec{x} \bullet \vec{y})\end{aligned} \tag{4.11.E.1}$$

But since U preserves distances, it is also the case that

$$(U(\vec{x}+\vec{y}) \bullet U(\vec{x}+\vec{y})) = ||\vec{x}||^2 + ||\vec{y}||^2 + 2(\vec{x} \bullet \vec{y})$$

Hence

$$\vec{x} \bullet \vec{y} = U^TU\vec{x} \bullet \vec{y}$$

and so

$$((U^TU - I)\vec{x}) \bullet \vec{y} = 0$$

Since $\vec{y}$ is arbitrary, it follows that $U^TU - I = 0$. Thus U is orthogonal.

Exercise $4.11.E.8$

You could observe that $\det(UU^T) = (\det(U))^2 - 1$ so $\det(U) \neq 0$.

Exercise $4.11.E.9$

$$\begin{aligned}&\begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & a \\ 0 & \frac{\sqrt{6}}{3} & b \end{bmatrix} \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & a \\ 0 & \frac{\sqrt{6}}{3} & b \end{bmatrix}^T \\ &= \begin{bmatrix} 1 & \frac{1}{3}\sqrt{3}a - \frac{1}{3} & \frac{1}{3}\sqrt{3}b - \frac{1}{3} \\ \frac{1}{3}\sqrt{3}a - \frac{1}{3} & a^2 + \frac{2}{3} & ab - \frac{1}{3} \\ \frac{1}{3}\sqrt{3}b - \frac{1}{3} & ab - \frac{1}{3} & b^2 + \frac{2}{3} \end{bmatrix}\end{aligned} \tag{4.11.E.2}$$

This requires, $a = 1/\sqrt{3}, b = 1/\sqrt{3}$.

$$\begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & 1/\sqrt{3} \\ 0 & \frac{\sqrt{6}}{3} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & 1/\sqrt{3} \\ 0 & \frac{\sqrt{6}}{3} & 1/\sqrt{3} \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Exercise 4.11.E.10

$$\begin{bmatrix} \frac{2}{3} & \frac{\sqrt{2}}{2} & \frac{1}{6}\sqrt{2} \\ \frac{2}{3} & \frac{-\sqrt{2}}{2} & a \\ -\frac{1}{3} & 0 & b \end{bmatrix} \begin{bmatrix} \frac{2}{3} & \frac{\sqrt{2}}{2} & \frac{1}{6}\sqrt{2} \\ \frac{2}{3} & \frac{-\sqrt{2}}{2} & a \\ -\frac{1}{3} & 0 & b \end{bmatrix}^T = \begin{bmatrix} 1 & \frac{1}{6}\sqrt{2}a - \frac{1}{18} & \frac{1}{6}\sqrt{2}b - \frac{2}{9} \\ \frac{1}{6}\sqrt{2}a - \frac{1}{18} & a^2 + \frac{17}{18} & ab - \frac{2}{9} \\ \frac{1}{6}\sqrt{2}b - \frac{2}{9} & ab - \frac{2}{9} & b^2 + \frac{1}{9} \end{bmatrix}$$

This requires $a = \frac{1}{3\sqrt{2}}$, $b = \frac{4}{3\sqrt{2}}$.

$$\begin{bmatrix} \frac{2}{3} & \frac{\sqrt{2}}{2} & \frac{1}{6}\sqrt{2} \\ \frac{2}{3} & \frac{-\sqrt{2}}{2} & \frac{1}{3\sqrt{2}} \\ -\frac{1}{3} & 0 & \frac{4}{3\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{2}{3} & \frac{\sqrt{2}}{2} & \frac{1}{6}\sqrt{2} \\ \frac{2}{3} & \frac{-\sqrt{2}}{2} & \frac{1}{3\sqrt{2}} \\ -\frac{1}{3} & 0 & \frac{4}{3\sqrt{2}} \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Exercise 4.11.E.11

Try

$$\begin{aligned} &\begin{bmatrix} \frac{1}{3} & -\frac{2}{\sqrt{5}} & c \\ \frac{2}{3} & 0 & d \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{15}\sqrt{5} \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{2}{\sqrt{5}} & c \\ \frac{2}{3} & 0 & d \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{15}\sqrt{5} \end{bmatrix}^T \\ &= \begin{bmatrix} c^2 + \frac{41}{45} & cd + \frac{2}{9} & \frac{4}{15}\sqrt{5}c - \frac{8}{45} \\ cd + \frac{2}{9} & d^2 + \frac{4}{9} & \frac{4}{15}\sqrt{5}d + \frac{4}{9} \\ \frac{4}{15}\sqrt{5}c - \frac{8}{45} & \frac{4}{15}\sqrt{5}d + \frac{4}{9} & 1 \end{bmatrix} \end{aligned} \tag{4.11.E.3}$$

This requires that $c = \frac{2}{3\sqrt{5}}, d = \frac{-5}{3\sqrt{5}}$.

$$\begin{bmatrix} \frac{1}{3} & -\frac{2}{\sqrt{5}} & \frac{2}{3\sqrt{5}} \\ \frac{2}{3} & 0 & \frac{-5}{3\sqrt{5}} \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{15}\sqrt{5} \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{2}{\sqrt{5}} & \frac{2}{3\sqrt{5}} \\ \frac{2}{3} & 0 & \frac{-5}{3\sqrt{5}} \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{15}\sqrt{5} \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Exercise 4.11.E.14

Compute the (i, j) entry of $A^T A$ explicitly.

Exercise 4.11.E.15

Express the inner product using matrix multiplication and use the fact that $Q^T Q = I$.

This page titled 4.11.E: Exercise for Section 4.8 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

4.12: Gram-Schmidt Process

Learning Objectives

- Given a linearly independent set, use the Gram-Schmidt Process to find corresponding orthogonal and orthonormal sets.

The Gram-Schmidt process is an algorithm to transform a set of vectors into an orthonormal set spanning the same subspace, that is generating the same collection of linear combinations.

The goal of the Gram-Schmidt process is to take a linearly independent set of vectors and transform it into an orthonormal set with the same span. The first objective is to construct an orthogonal set of vectors with the same span, since from there an orthonormal set can be obtained by simply dividing each vector by its length.

Algorithm 4.12.1: Gram-Schmidt Process

Let $\{\vec{u}_1, \cdots, \vec{u}_n\}$ be a set of linearly independent vectors in $\mathbb{R}^n$.

I: Construct a new set of vectors $\{\vec{v}_1, \cdots, \vec{v}_n\}$ as follows:

$$\begin{aligned}
\vec{v}_1 &= \vec{u}_1 \\
\vec{v}_2 &= \vec{u}_2 - \left(\frac{\vec{u}_2 \cdot \vec{v}_1}{\|\vec{v}_1\|^2}\right)\vec{v}_1 \\
\vec{v}_3 &= \vec{u}_3 - \left(\frac{\vec{u}_3 \cdot \vec{v}_1}{\|\vec{v}_1\|^2}\right)\vec{v}_1 - \left(\frac{\vec{u}_3 \cdot \vec{v}_2}{\|\vec{v}_2\|^2}\right)\vec{v}_2 \\
&\vdots \\
\vec{v}_n &= \vec{u}_n - \left(\frac{\vec{u}_n \cdot \vec{v}_1}{\|\vec{v}_1\|^2}\right)\vec{v}_1 - \left(\frac{\vec{u}_n \cdot \vec{v}_2}{\|\vec{v}_2\|^2}\right)\vec{v}_2 - \cdots - \left(\frac{\vec{u}_n \cdot \vec{v}_{n-1}}{\|\vec{v}_{n-1}\|^2}\right)\vec{v}_{n-1}
\end{aligned}$$

II: Now let $\vec{w}_i = \dfrac{\vec{v}_i}{\|\vec{v}_i\|}$ for $i = 1, \cdots, n$.

Then

1. $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is an orthogonal set.
2. $\{\vec{w}_1, \cdots, \vec{w}_n\}$ is an orthonormal set.
3. $\text{span}\{\vec{u}_1, \cdots, \vec{u}_n\} = \text{span}\{\vec{v}_1, \cdots, \vec{v}_n\} = \text{span}\{\vec{w}_1, \cdots, \vec{w}_n\}$.

Solution

The full proof of this algorithm is beyond this material, however here is an indication of the arguments.

To show that $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is an orthogonal set, let

$$a_2 = \frac{\vec{u}_2 \cdot \vec{v}_1}{\|\vec{v}_1\|^2}$$

then:

$$\begin{aligned}
\vec{v}_1 \cdot \vec{v}_2 &= \vec{v}_1 \cdot (\vec{u}_2 - a_2\vec{v}_1) \\
&= \vec{v}_1 \cdot \vec{u}_2 - a_2(\vec{v}_1 \cdot \vec{v}_1) \\
&= \vec{v}_1 \cdot \vec{u}_2 - \frac{\vec{u}_2 \cdot \vec{v}_1}{\|\vec{v}_1\|^2}\|\vec{v}_1\|^2 \\
&= (\vec{v}_1 \cdot \vec{u}_2) - (\vec{u}_2 \cdot \vec{v}_1) = 0
\end{aligned}$$

Now that you have shown that $\{\vec{v}_1, \vec{v}_2\}$ is orthogonal, use the same method as above to show that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is also orthogonal, and so on.

Then in a similar fashion you show that $\text{span}\{\vec{u}_1, \cdots, \vec{u}_n\} = \text{span}\{\vec{v}_1, \cdots, \vec{v}_n\}$.

Finally defining $\vec{w}_i = \dfrac{\vec{v}_i}{\|\vec{v}_i\|}$ for $i = 1, \cdots, n$ does not affect orthogonality and yields vectors of length 1, hence an orthonormal set. You can also observe that it does not affect the span either and the proof would be complete.

Consider the following example.

Example 4.12.1: Find Orthonormal Set with Same Span

Consider the set of vectors $\{\vec{u}_1, \vec{u}_2\}$ given as in Example 4.12.1. That is

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} \overset{\text{int}}{\mathbb{R}^3}$$

Use the Gram-Schmidt algorithm to find an orthonormal set of vectors $\{\vec{w}_1, \vec{w}_2\}$ having the same span.

Solution

We already remarked that the set of vectors in $\{\vec{u}_1, \vec{u}_2\}$ is linearly independent, so we can proceed with the Gram-Schmidt algorithm:

$$\begin{aligned} \vec{v}_1 = \vec{u}_1 &= \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \\ \vec{v}_2 = \vec{u}_2 - \left(\frac{\vec{u}_2 \cdot \vec{v}_1}{\|\vec{v}_1\|^2} \right) \vec{v}_1 & \\ &= \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} - \frac{5}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 0 \end{bmatrix} \end{aligned} \tag{4.12.1}$$

Now to normalize simply let

$$\begin{aligned} \vec{w}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} &= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \\ \vec{w}_2 = \frac{\vec{v}_2}{\|\vec{v}_2\|} &= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \end{aligned} \tag{4.12.2}$$

You can verify that $\{\vec{w}_1, \vec{w}_2\}$ is an orthonormal set of vectors having the same span as $\{\vec{u}_1, \vec{u}_2\}$, namely the XY-plane.

In this example, we began with a linearly independent set and found an orthonormal set of vectors which had the same span. It turns out that if we start with a basis of a subspace and apply the Gram-Schmidt algorithm, the result will be an orthogonal basis of the same subspace. We examine this in the following example.

Example 4.12.2: Find a Corresponding Orthogonal Basis

Let

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \vec{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \text{ and } \vec{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

and let $U = \text{span}\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$. Use the Gram-Schmidt Process to construct an orthogonal basis B of U.

Solution

First $\vec{f}_1 = \vec{x}_1$.

Next,

$$\vec{f}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Finally,

$$\vec{f}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \frac{0}{1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1 \\ -1/2 \\ 0 \end{bmatrix}.$$

Therefore,

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 1 \\ -1/2 \\ 0 \end{bmatrix} \right\}$$

is an orthogonal basis of U. However, it is sometimes more convenient to deal with vectors having integer entries, in which case we take

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix} \right\}.$$

This page titled 4.12: Gram-Schmidt Process is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

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

4.12.E: Exercises for Section 4.9

Exercise 4.12.E.1

Apply the Gram-Schmidt process to the following set of vectors in $\mathbb{R}^2$ to create an orthonormal basis:

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Answer available at the end of this section.

Exercise 4.12.E.2

Find an orthonormal basis for the span of each of the following sets of vectors.

a. $\begin{bmatrix} 3 \\ -4 \\ 0 \end{bmatrix}, \begin{bmatrix} 7 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 7 \\ 1 \end{bmatrix}$

b. $\begin{bmatrix} 3 \\ 0 \\ -4 \end{bmatrix}, \begin{bmatrix} 11 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 7 \end{bmatrix}$

c. $\begin{bmatrix} 3 \\ 0 \\ -4 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \\ 10 \end{bmatrix}, \begin{bmatrix} -7 \\ 1 \\ 1 \end{bmatrix}$

Answer available at the end of this section.

Exercise 4.12.E.3

Using the Gram Schmidt process find an orthonormal basis for the following span:

$$\text{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

Answer available at the end of this section.

Exercise 4.12.E.4

Using the Gram Schmidt process find an orthonormal basis for the following span:

$$\text{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Answer available at the end of this section.

Exercise 4.12.E.5

Consider the following vectors:

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} a \\ b \end{bmatrix} \tag{4.12.E.1}$$

a. Find an orthonormal basis.
b. What restriction did you have to make for your answer to part a?

Answer available at the end of this section.

Exercise 4.12.E.6

The set $V = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : 2x + 3y - z = 0 \right\}$ is a subspace of $\mathbb{R}^3$. Find an orthonormal basis for this subspace.

Answer available at the end of this section.

Exercise 4.12.E.7

Use the Gram Schmidt process to transform the given basis for a subspace in $\mathbb{R}^n$ into an orthogonal basis for the subspace.

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

b. $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

Answers

Exercise 4.12.E.1

$$\left\{ \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$$

Exercise 4.12.E.2

a. $\begin{bmatrix} \frac{3}{5} \\ -\frac{4}{5} \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{4}{5} \\ \frac{3}{5} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

b. $\begin{bmatrix} \frac{3}{5} \\ 0 \\ -\frac{4}{5} \end{bmatrix}, \begin{bmatrix} \frac{4}{5} \\ 0 \\ \frac{3}{5} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

c. $\begin{bmatrix} \frac{3}{5} \\ 0 \\ -\frac{4}{5} \end{bmatrix}, \begin{bmatrix} \frac{4}{5} \\ 0 \\ \frac{3}{5} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

Exercise 4.12.E.3

A solution is

$$\begin{bmatrix} \frac{1}{6}\sqrt{6} \\ \frac{1}{3}\sqrt{6} \\ \frac{1}{6}\sqrt{6} \end{bmatrix}, \begin{bmatrix} \frac{3}{10}\sqrt{2} \\ -\frac{2}{5}\sqrt{2} \\ \frac{1}{2}\sqrt{2} \end{bmatrix}, \begin{bmatrix} \frac{7}{15}\sqrt{3} \\ -\frac{1}{15}\sqrt{3} \\ -\frac{1}{3}\sqrt{3} \end{bmatrix}$$

Exercise 4.12.E.4

Then a solution is

$$\left[\begin{array}{c} \frac{1}{6}\sqrt{6} \\ \frac{1}{3}\sqrt{6} \\ \frac{1}{6}\sqrt{6} \\ 0 \end{array}\right], \left[\begin{array}{c} \frac{1}{6}\sqrt{2}\sqrt{3} \\ -\frac{2}{9}\sqrt{2}\sqrt{3} \\ \frac{5}{18}\sqrt{2}\sqrt{3} \\ \frac{1}{9}\sqrt{2}\sqrt{3} \end{array}\right], \left[\begin{array}{c} \frac{5}{111}\sqrt{3}\sqrt{37} \\ \frac{1}{133}\sqrt{3}\sqrt{37} \\ -\frac{17}{333}\sqrt{3}\sqrt{37} \\ \frac{22}{333}\sqrt{3}\sqrt{37} \end{array}\right]$$

Exercise $4.12.E.5$

a. $$\left\{\frac{1}{\sqrt{5}}\begin{bmatrix}1\\2\end{bmatrix}, \frac{1}{\sqrt{a^2+b^2-\frac{(a+2b)^2}{5}}}\begin{bmatrix}a-\frac{a+2b}{5}\\ b-\frac{2(a+2b)}{5}\end{bmatrix}\right\}$$

b. $b \neq 2a$.

Exercise $4.12.E.6$

The subspace is of the form

$$\left[\begin{array}{c} x \\ y \\ 2x+3y \end{array}\right]$$

and a basis is $\left[\begin{array}{c} 1 \\ 0 \\ 2 \end{array}\right], \left[\begin{array}{c} 0 \\ 1 \\ 3 \end{array}\right]$. Therefore, an orthonormal basis is

$$\left[\begin{array}{c} \frac{1}{5}\sqrt{5} \\ 0 \\ \frac{2}{5}\sqrt{5} \end{array}\right], \left[\begin{array}{c} -\frac{3}{35}\sqrt{5}\sqrt{14} \\ \frac{1}{14}\sqrt{5}\sqrt{14} \\ \frac{3}{70}\sqrt{5}\sqrt{14} \end{array}\right]$$

This page titled 4.12.E: Exercises for Section 4.9 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

4.13: 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 4.13.1: 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 4.13.1

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 $W^\perp$.

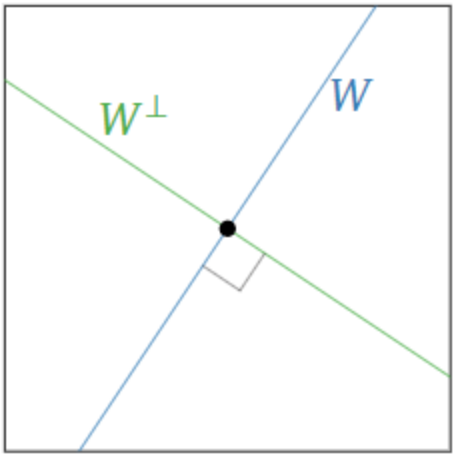


Figure 4.13.1

Example 4.13.1: Interactive: Orthogonal complements in $\mathbb{R}^2$

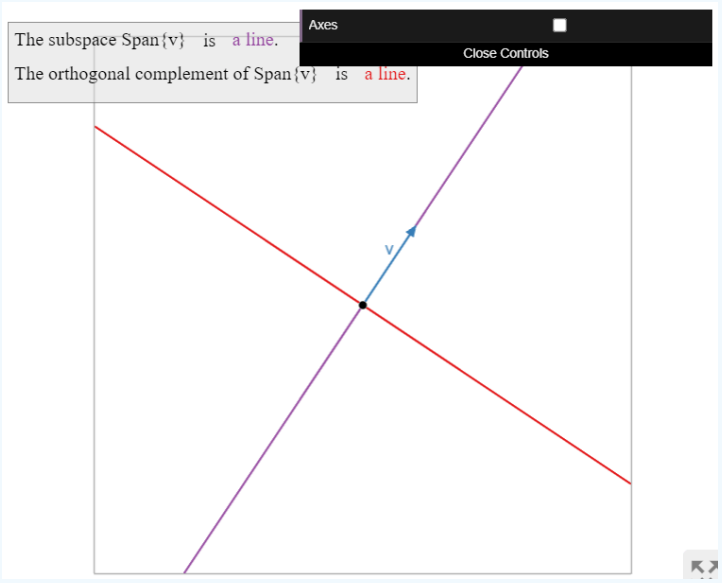


Figure 4.13.2: The orthogonal complement of the line spanned by v is the perpendicular line. Click and drag the head of v to move it.

The orthogonal complement of a line W in $\mathbb{R}^3$ is the perpendicular plane $W^\perp$.

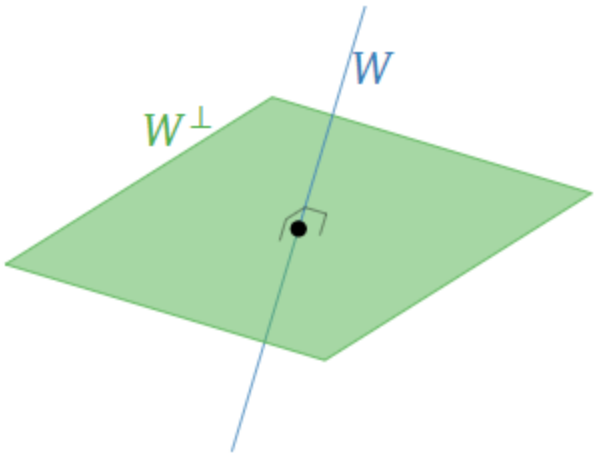


Figure 4.13.3

Example 4.13.2: Interactive: Orthogonal complements in $\mathbb{R}^3$

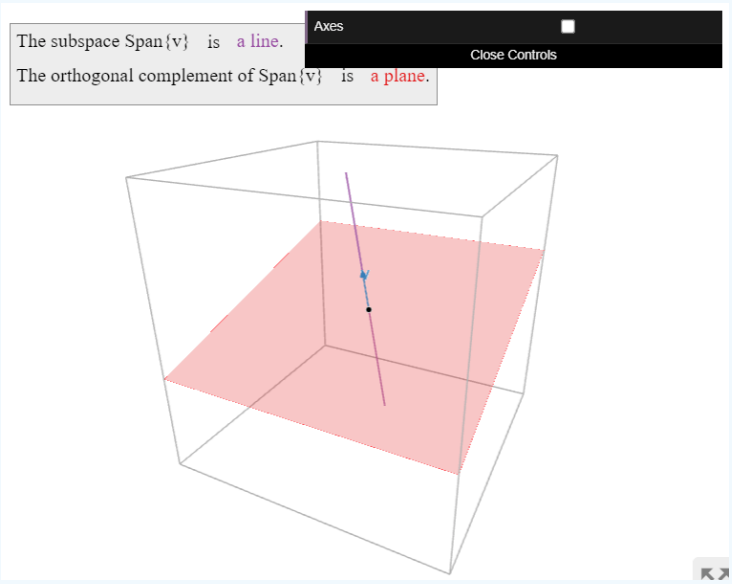


Figure 4.13.4: The orthogonal complement of the line spanned by v is the perpendicular plane. Click and drag the head of v to move it.

The orthogonal complement of a plane W in $\mathbb{R}^3$ is the perpendicular line $W^\perp$.

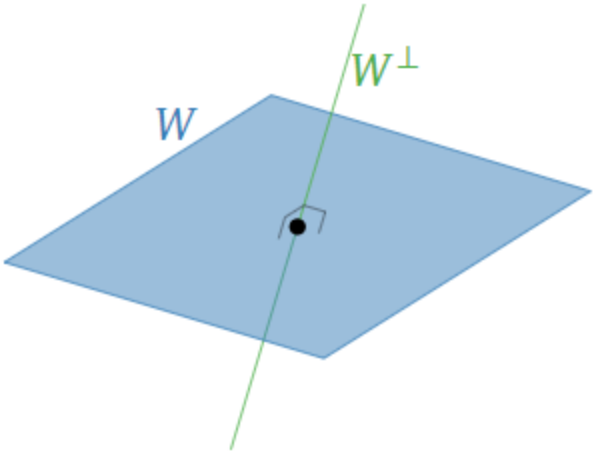


Figure 4.13.5

Example 4.13.3: Interactive: Orthogonal complements in $\mathbb{R}^3$

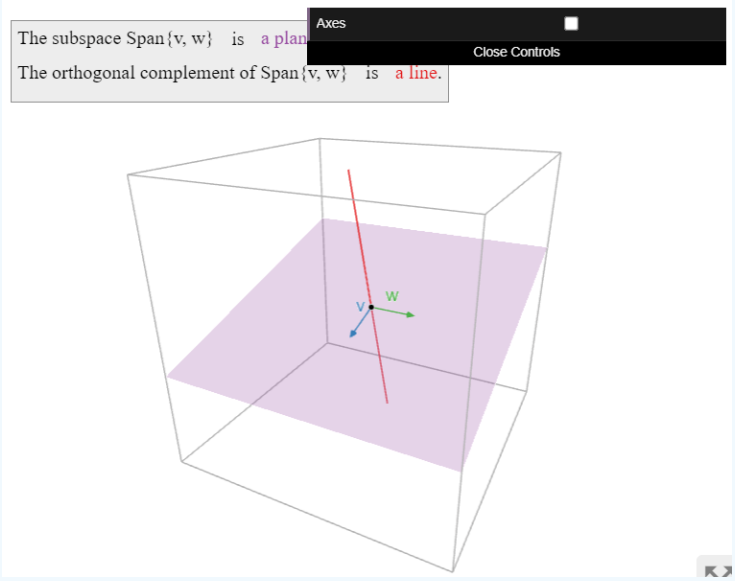


Figure 4.13.6: The orthogonal complement of the plane spanned by v, w is the perpendicular line. Click and drag the heads of v, w to change the plane.

We see in the above pictures that $(W^\perp)^\perp = W$.

Example 4.13.4

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 4.13.1: The Orthogonal Complement of a Column Space

Let A be a matrix and let $W = \mathrm{Col}(A)$. Then

$$W^\perp = \mathrm{Nul}(A^T).$$

Proof available at the end of this section.

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 $W = \mathrm{Span}\{v_1, v_2, \ldots, v_m\}$. Then

$$W^\perp = \{\text{all vectors orthogonal to each } v_1, v_2, \ldots, v_m\} = \mathrm{Nul}\begin{pmatrix} —v_1^T— \\ —v_2^T— \\ \vdots \\ —v_m^T— \end{pmatrix}.$$

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 $\mathrm{Span}\{v_1, v_2\}^\perp$ 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 4.13.5

Compute $W^\perp$, where

$$W = \mathrm{Span}\left\{\begin{pmatrix} 1 \\ 7 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 3 \\ 1 \end{pmatrix}\right\}.$$

Solution

According to Proposition 4.13.1, 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

$$W^\perp = \text{Span}\left\{ \begin{pmatrix} 1 \\ -5 \\ 17 \end{pmatrix} \right\}.$$

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 4.13.6

Find all vectors orthogonal to $v = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$.

Solution

According to Proposition $4.13.1$, we need to compute the null space of the matrix

$$A = \begin{pmatrix} \text{—} v \text{—} \end{pmatrix} = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}.$$

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*

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

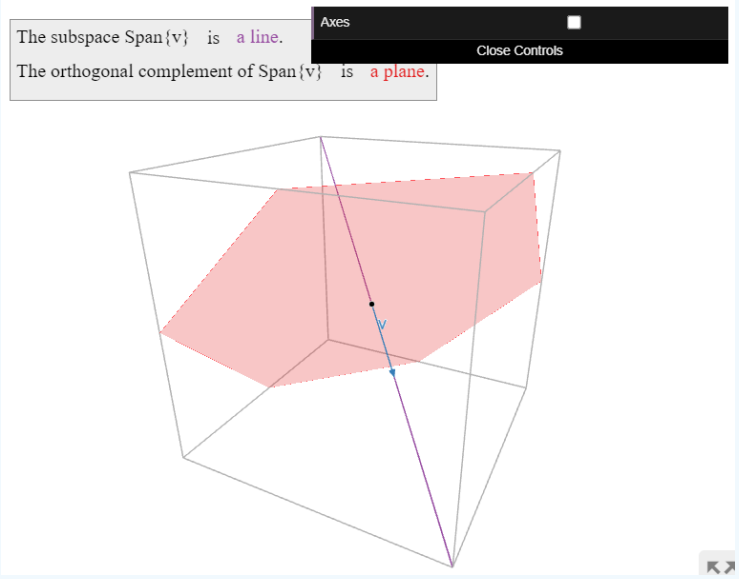


Figure 4.13.7: The set of all vectors perpendicular to v.

Example 4.13.7

Compute

$$\text{Span}\left\{\begin{pmatrix}1\\1\\-1\end{pmatrix},\begin{pmatrix}1\\1\\1\end{pmatrix}\right\}^{\perp}.$$

Solution

According to Proposition 4.13.1, 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*

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

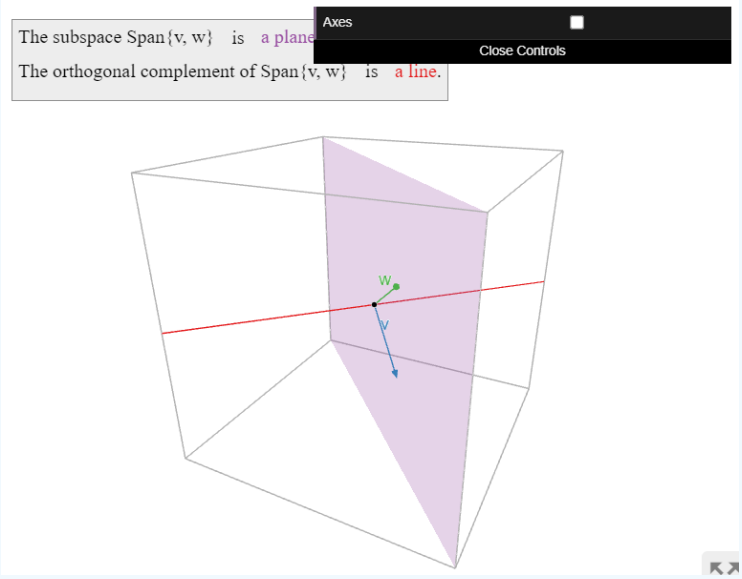


Figure 4.13.8: The orthogonal complement of the plane spanned by $v = (1, 1, -1)$ and $w = (1, 1, 1)$.

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 4.13.1: 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 available at the end of this section.

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

$$\begin{aligned} \text{Span}\{e_1, e_2\}^\perp &= \left\{ \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \text{ in } \mathbb{R} \;\middle|\; \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = 0 \text{ and } \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = 0 \right\} \\ &= \left\{ \begin{pmatrix} 0 \\ 0 \\ z \\ w \end{pmatrix} \text{ in } \mathbb{R}^4 \right\} = \text{Span}\{e_3, e_4\} : \end{aligned} \tag{4.13.1}$$

the orthogonal complement of the xy-plane is the zw-plane.

Definition 4.13.2: 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 4.13.3 that if A has rows $v_1^T, v_2^T, \dots, v_m^T$, then

$$\text{Row}(A)^\perp = \text{Span}\{v_1, v_2, \dots, v_m\}^\perp = \text{Nul}(A).$$

Taking orthogonal complements of both sides and using the second fact 4.13.1 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 $v_1, v_2, \dots, v_m$, we have

$$\text{Span}\{v_1, v_2, \dots, v_m\}^\perp = \text{Nul}\begin{pmatrix} -v_1^T- \\ -v_2^T- \\ \vdots \\ -v_m^T- \end{pmatrix}.$$

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 4.13.8: 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 4.13.9: 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 $\text{Col}(A)$ and $\text{Row}(A)$ 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, $\text{Row}(A)$ lies in $\mathbb{R}^n$ and $\text{Col}(A)$ 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 4.13.1

Let A be a matrix. Then the row rank of A is equal to the column rank of A.

Proof available at the end of this section.

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.

Answers

Proposition 4.13.1: The Orthogonal Complement of a Column Space

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

$$A = \begin{pmatrix} \text{—} v_1^T \text{—} \\ \text{—} v_2^T \text{—} \\ \vdots \\ \text{—} v_m^T \text{—} \end{pmatrix}.$$

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 $\mathrm{Nul}(A)$ if and only if x is perpendicular to each vector $v_1, v_2, \ldots, v_m$.

Fact $4.13.1$**: Facts about Orthogonal Complements**

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 $\mathrm{Nul}(A)$. Then
$$0 = Ax = \begin{pmatrix} v_1^Tx \\ v_2^Tx \\ \vdots \\ v_k^Tx \end{pmatrix} = \begin{pmatrix} v_1 \cdot x \\ v_2 \cdot x \\ \vdots \\ v_k \cdot x \end{pmatrix}$$
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 $4.13.1$ 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 $(W^\perp)^\perp = W$.

Theorem $4.13.1$

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 $4.13.1$ 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.

This page titled 4.13: Orthogonal Complements is shared under a GNU FDL license and was authored, remixed, and/or curated by Dan Margalit and Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- 6.2: Orthogonal Complements by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

4.14: Orthogonal Projections

Learning Objectives

- Find the orthogonal projection of a vector onto a subspace.
- Find the orthogonal complement of a subspace.

In this section, we compute orthogonal projections of vectors onto subspaces.

An important use of the Gram-Schmidt Process is in orthogonal projections, the focus of this section.

You may recall that a subspace of $\mathbb{R}^n$ is a set of vectors which contains the zero vector, and is closed under addition and scalar multiplication. Let's call such a subspace W. In particular, a plane in $\mathbb{R}^n$ which contains the origin, $(0, 0, \cdots, 0)$, is a subspace of $\mathbb{R}^n$.

Suppose a point Y in $\mathbb{R}^n$ is not contained in W, then what point Z in W is closest to Y? Using the Gram-Schmidt Process, we can find such a point. Let $\vec{y}, \vec{z}$ represent the position vectors of the points Y and Z respectively, with $\vec{y} - \vec{z}$ representing the vector connecting the two points Y and Z. It will follow that if Z is the point on W closest to Y, then $\vec{y} - \vec{z}$ will be perpendicular to W (can you see why?); in other words, $\vec{y} - \vec{z}$ is orthogonal to W (and to every vector contained in W) as in the following diagram.

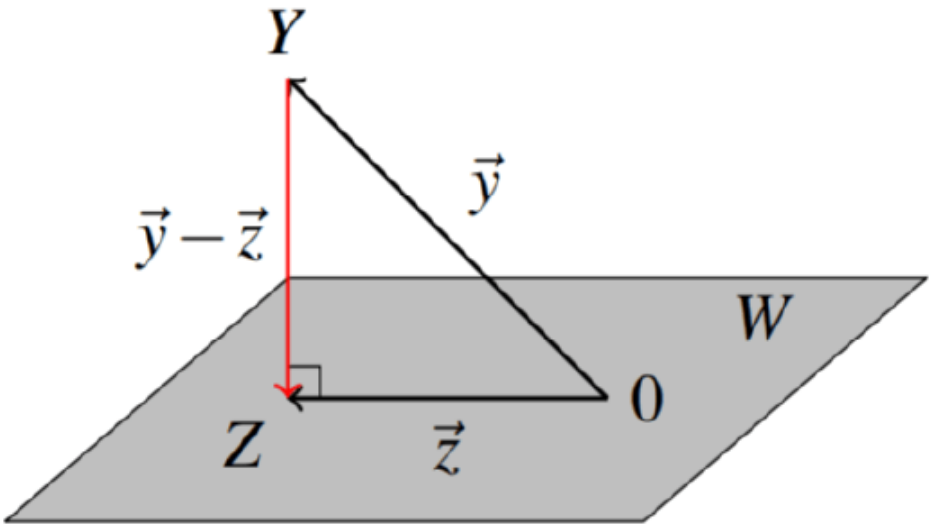


Figure 4.14.1 This diagram illustrates the concept of orthogonal projection of a vector onto a subspace. It shows a vector $\vec{y}$, a subspace W, and the orthogonal projection of $\vec{y}$ onto W, denoted as $\vec{z}$. The diagram helps visualize the relationship between the vector, the subspace, and the projection. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

The vector $\vec{z}$ is called the **orthogonal projection** of $\vec{y}$ on W. The definition is given as follows.

Definition 4.14.1: Orthogonal Projection

Let W be a subspace of $\mathbb{R}^n$, and Y be any point in $\mathbb{R}^n$. Then the **orthogonal projection** of Y onto W is given by

$$\vec{z} = \text{proj}_W(\vec{y}) = \left(\frac{\vec{y} \cdot \vec{w}_1}{\|\vec{w}_1\|^2}\right) \vec{w}_1 + \left(\frac{\vec{y} \cdot \vec{w}_2}{\|\vec{w}_2\|^2}\right) \vec{w}_2 + \cdots + \left(\frac{\vec{y} \cdot \vec{w}_m}{\|\vec{w}_m\|^2}\right) \vec{w}_m$$

where $\{\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_m\}$ is any orthogonal basis of W.

Therefore, in order to find the orthogonal projection, we must first find an orthogonal basis for the subspace. Note that one could use an orthonormal basis, but it is not necessary in this case since as you can see above the normalization of each vector is included in the formula for the projection.

Before we explore this further through an example, we show that the orthogonal projection does indeed yield a point Z (the point whose position vector is the vector $\vec{z}$ above) which is the point of W closest to Y.

Theorem 4.14.1: Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$ and Y any point in $\mathbb{R}^n$. Let Z be the point whose position vector is the orthogonal projection of Y onto W.

Then, Z is the point in W closest to Y.

Proof available at the end of this section.

Consider the following example.

Example 4.14.1: Orthogonal Projection

Let W be the plane through the origin given by the equation $x - 2y + z = 0$.
Find the point in W closest to the point $Y = (1, 0, 3)$.

Solution

We must first find an orthogonal basis for W. Notice that W is characterized by all points (a, b, c) where $c = 2b - a$. In other words,

$$W = \begin{bmatrix} a \\ b \\ 2b - a \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \ a, b^{\text{int}}\mathbb{R}$$

We can thus write W as

$$\begin{aligned} W &= \text{span}\left\{\vec{u}_1, \vec{u}_2\right\} \\ &= \text{span}\left\{\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}\right\} \end{aligned} \tag{4.14.1}$$

Notice that this span is a basis of W as it is linearly independent. We will use the Gram-Schmidt Process to convert this to an orthogonal basis, $\left\{\vec{w}_1, \vec{w}_2\right\}$. In this case, as we remarked it is only necessary to find an orthogonal basis, and it is not required that it be orthonormal.

$$\vec{w}_1 = \vec{u}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

$$\begin{aligned}\vec{w}_2 &= \vec{u}_2 - \left(\frac{\vec{u}_2 \cdot \vec{w}_1}{\|\vec{w}_1\|^2}\right)\vec{w}_1 \\ &= \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} - \left(\frac{-2}{2}\right)\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\end{aligned} \tag{4.14.2}$$

Therefore an orthogonal basis of W is

$$\{\vec{w}_1, \vec{w}_2\} = \left\{\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right\}$$

We can now use this basis to find the orthogonal projection of the point $Y = (1, 0, 3)$ on the subspace W. We will write the position vector $\vec{y}$ of Y as $\vec{y} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$. We compute the projection as follows:

$$\begin{aligned}\vec{z} &= \text{proj}_W(\vec{y}) \\ &= \left(\frac{\vec{y} \cdot \vec{w}_1}{\|\vec{w}_1\|^2}\right)\vec{w}_1 + \left(\frac{\vec{y} \cdot \vec{w}_2}{\|\vec{w}_2\|^2}\right)\vec{w}_2 \\ &= \left(\frac{-2}{2}\right)\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \left(\frac{4}{3}\right)\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \\ \frac{7}{3} \end{bmatrix}\end{aligned} \tag{4.14.3}$$

Therefore the point Z on W closest to the point $(1, 0, 3)$ is $(\frac{1}{3}, \frac{4}{3}, \frac{7}{3})$.

Recall that the vector $\vec{y} - \vec{z}$ is perpendicular (orthogonal) to all the vectors contained in the plane W. Using a basis for W, we can in fact find all such vectors which are perpendicular to W. We call this set of vectors the **orthogonal complement** of W and denote it $W^\perp$.

Definition 4.14.2: Orthogonal Complement

Let W be a subspace of $\mathbb{R}^n$. Then the **orthogonal complement** of W, written $W^\perp$, is the set of all vectors $\vec{x}$ such that $\vec{x} \cdot \vec{z} = 0$ for all vectors $\vec{z}$ in W.

$$W^\perp = \{\vec{x} \in \mathbb{R}^n \text{ such that } \vec{x} \cdot \vec{z} = 0 \text{ for all } \vec{z} \in W\}$$

The orthogonal complement is defined as the set of all vectors which are orthogonal to all vectors in the original subspace. It turns out that it is sufficient that the vectors in the orthogonal complement be orthogonal to a spanning set of the original space.

Proposition 4.14.1: Orthogonal to Spanning Set

Let W be a subspace of $\mathbb{R}^n$ such that $W = \text{span}\{\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_m\}$. Then $W^\perp$ is the set of all vectors which are orthogonal to each $\vec{w}_i$ in the spanning set.

The following proposition demonstrates that the orthogonal complement of a subspace is itself a subspace.

Proposition 4.14.2: The Orthogonal Complement

Let W be a subspace of $\mathbb{R}^n$. Then the orthogonal complement $W^\perp$ is also a subspace of $\mathbb{R}^n$.

Consider the following proposition.

Proposition 4.14.3: Orthogonal Complement of $\mathbb{R}^n$

The complement of $\mathbb{R}^n$ is the set containing the zero vector:

$$(\mathbb{R}^n)^\perp = \left\{\vec{0}\right\}$$

Similarly,

$$\left\{\vec{0}\right\}^\perp = (\mathbb{R}^n).$$

Proof available at the end of this section.

In the next example, we will look at how to find $W^\perp$.

Example 4.14.2: Orthogonal Complement

Let W be the plane through the origin given by the equation $x - 2y + z = 0$. Find a basis for the orthogonal complement of W.

Solution

From Example 4.14.1 we know that we can write W as

$$W = \text{span}\left\{\vec{u}_1, \vec{u}_2\right\} = \text{span}\left\{\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}\right\}$$

In order to find $W^\perp$, we need to find all $\vec{x}$ which are orthogonal to every vector in this span.

Let $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$. In order to satisfy $\vec{x} \cdot \vec{u}_1 = 0$, the following equation must hold.

$$x_1 - x_3 = 0$$

In order to satisfy $\vec{x} \cdot \vec{u}_2 = 0$, the following equation must hold.

$$x_2 + 2x_3 = 0$$

Both of these equations must be satisfied, so we have the following system of equations.

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

To solve, set up the augmented matrix.

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

Using Gaussian Elimination, we find that $W^{\perp} = \text{span}\left\{\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}\right\}$, and hence $\left\{\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}\right\}$ is a basis for $W^{\perp}$.

The following results summarize the important properties of the orthogonal projection.

Theorem 4.14.2: Orthogonal Projection

Let W be a subspace of $\mathbb{R}^n$, Y be any point in $\mathbb{R}^n$, and let Z be the point in W closest to Y. Then,

1. The position vector $\vec{z}$ of the point Z is given by $\vec{z} = \text{proj}_W(\vec{y})$
2. $\vec{z}\,^{\text{int}}W$ and $\vec{y} - \vec{z}\,^{\text{int}}W^{\perp}$
3. $|Y - Z| < |Y - Z_1|$ for all $Z_1 \neq Z^{\text{int}}W$

Consider the following example of this concept.

Example 4.14.3: Find a Vector Closest to a Given Vector

Let

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \vec{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \vec{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \text{ and } \vec{v} = \begin{bmatrix} 4 \\ 3 \\ -2 \\ 5 \end{bmatrix}.$$

We want to find the vector in $W = \text{span}\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$ closest to $\vec{y}$.

Solution

We will first use the Gram-Schmidt Process to construct the orthogonal basis, B, of W:

$$B = \left\{\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix}\right\}.$$

By Theorem 4.14.2,

$$\text{proj}_U(\vec{v}) = \frac{2}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \frac{5}{1}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \frac{12}{6}\begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ -1 \\ 5 \end{bmatrix}$$

is the vector in U closest to $\vec{y}$.

Consider the next example.

Example 4.14.4: Vector Written as a Sum of Two Vectors

Let W be a subspace given by $W = \text{span}\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 2 \end{bmatrix}\right\}$, and $Y = (1, 2, 3, 4)$.

Find the point Z in W closest to Y, and moreover write $\vec{y}$ as the sum of a vector in W and a vector in $W^{\perp}$.

Solution

From Theorem 4.14.3 the point Z in W closest to Y is given by $\vec{z} = \text{proj}_W(\vec{y})$.

Notice that since the above vectors already give an orthogonal basis for W, we have:

$$\begin{aligned} \vec{z} &= \text{proj}_W(\vec{y}) \\ &= \left(\frac{\vec{y}\cdot\vec{w}_1}{\|\vec{w}_1\|^2}\right)\vec{w}_1 + \left(\frac{\vec{y}\cdot\vec{w}_2}{\|\vec{w}_2\|^2}\right)\vec{w}_2 \\ &= \left(\frac{4}{2}\right)\begin{bmatrix}1\\0\\1\\0\end{bmatrix} + \left(\frac{10}{5}\right)\begin{bmatrix}0\\1\\0\\2\end{bmatrix} \\ &= \begin{bmatrix}2\\2\\2\\4\end{bmatrix} \end{aligned} \tag{4.14.4}$$

Therefore the point in W closest to Y is $Z = (2,2,2,4)$.
Now, we need to write $\vec{y}$ as the sum of a vector in W and a vector in $W^\perp$. This can easily be done as follows:

$$\vec{y} = \vec{z} + (\vec{y} - \vec{z})$$

since $\vec{z}$ is in W and as we have seen $\vec{y} - \vec{z}$ is in $W^\perp$.
The vector $\vec{y} - \vec{z}$ is given by

$$\vec{y} - \vec{z} = \begin{bmatrix}1\\2\\3\\4\end{bmatrix} - \begin{bmatrix}2\\2\\2\\4\end{bmatrix} = \begin{bmatrix}-1\\0\\1\\0\end{bmatrix}$$

Therefore, we can write $\vec{y}$ as

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

Example 4.14.5: Point in a Plane Closest to a Given Point

Find the point Z in the plane $3x + y - 2z = 0$ that is closest to the point $Y = (1,1,1)$.

Solution

The solution will proceed as follows.

1. Find a basis X of the subspace W of $\mathbb{R}^3$ defined by the equation $3x + y - 2z = 0$.
2. Orthogonalize the basis X to get an orthogonal basis B of W.
3. Find the projection on W of the position vector of the point Y.

We now begin the solution.

1. $3x + y - 2z = 0$ is a system of one equation in three variables. Putting the augmented matrix in reduced row-echelon form:

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

   gives general solution $x = \frac{1}{3}s + \frac{2}{3}t$, $y = s$, $z = t$ for any $s, t \in \mathbb{R}$. Then

$$W = \text{span}\left\{\begin{bmatrix} -\frac{1}{3} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{2}{3} \\ 0 \\ 1 \end{bmatrix}\right\}$$

Let $X = \left\{\begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix}\right\}$. Then X is linearly independent and $\text{span}(X) = W$, so X is a basis of W.

2. Use the Gram-Schmidt Process to get an orthogonal basis of W:

$$\vec{f}_1 = \begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix} \text{ and } \vec{f}_2 = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} - \frac{-2}{10}\begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 9 \\ 3 \\ 15 \end{bmatrix}.$$

Therefore $B = \left\{\begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}\right\}$ is an orthogonal basis of W.

3. To find the point Z on W closest to $Y = (1, 1, 1)$, compute

$$\begin{aligned} \text{proj}_W \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} &= \frac{2}{10}\begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix} + \frac{9}{35}\begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix} \\ &= \frac{1}{7}\begin{bmatrix} 4 \\ 6 \\ 9 \end{bmatrix}. \end{aligned} \tag{4.14.5}$$

Therefore, $Z = \left(\frac{4}{7}, \frac{6}{7}, \frac{9}{7}\right)$.

Answers

Theorem 4.14.1**: Approximation Theorem**

First Z is certainly a point in W since it is in the span of a basis of W.

To show that Z is the point in W closest to Y, we wish to show that $|\vec{y} - \vec{z}_1| > |\vec{y} - \vec{z}|$ for all $\vec{z}_1 \neq \vec{z}^{\,\text{int}} W$. We begin by writing $\vec{y} - \vec{z}_1 = (\vec{y} - \vec{z}) + (\vec{z} - \vec{z}_1)$. Now, the vector $\vec{y} - \vec{z}$ is orthogonal to W, and $\vec{z} - \vec{z}_1$ is contained in W. Therefore these vectors are orthogonal to each other. By the Pythagorean Theorem, we have that

$$\|\vec{y} - \vec{z}_1\|^2 = \|\vec{y} - \vec{z}\|^2 + \|\vec{z} - \vec{z}_1\|^2 > \|\vec{y} - \vec{z}\|^2$$

This follows because $\vec{z} \neq \vec{z}_1$ so $\|\vec{z} - \vec{z}_1\|^2 > 0$.

Hence, $\|\vec{y} - \vec{z}_1\|^2 > \|\vec{y} - \vec{z}\|^2$. Taking the square root of each side, we obtain the desired result.

Proposition 4.14.3**: Orthogonal Complement of** $\mathbb{R}^n$

Here, $\vec{0}$ is the zero vector of $\mathbb{R}^n$. Since $\vec{x} \cdot \vec{0} = 0$ for all $\vec{x}^{\,\text{int}} \mathbb{R}^n$, $\mathbb{R}^n \subseteq \{\vec{0}\}^\perp$. Since $\{\vec{0}\}^\perp \subseteq \mathbb{R}^n$, the equality follows, i.e., $\{\vec{0}\}^\perp = \mathbb{R}^n$.

Again, since $\vec{x} \cdot \vec{0} = 0$ for all $\vec{x}^{\,\text{int}} \mathbb{R}^n$, $\vec{0}^{\,\text{int}} (\mathbb{R}^n)^\perp$, so $\{\vec{0}\} \subseteq (\mathbb{R}^n)^\perp$. Suppose $\vec{x}^{\,\text{int}} \mathbb{R}^n$, $\vec{x} \neq \vec{0}$. Since $\vec{x} \cdot \vec{x} = ||\vec{x}||^2$ and $\vec{x} \neq \vec{0}$, $\vec{x} \cdot \vec{x} \neq 0$, so $\vec{x} \not{}^{\text{int}} (\mathbb{R}^n)^\perp$. Therefore $(\mathbb{R}^n)^\perp \subseteq \{\vec{0}\}$, and thus $(\mathbb{R}^n)^\perp = \{\vec{0}\}$.

This page titled 4.14: Orthogonal Projections is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

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

4.14.E: Exercise for Section 4.10

Exercise 4.14.E.1

Let the subspace W of $\mathbb{R}^3$ be spanned by the vectors:

$$\vec{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{w}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Find the orthogonal projection of the vector:

$$\vec{v} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

onto the subspace W.

Answer available at the end of this section.

Exercise 4.14.E.2

Let W be the subspace of $\mathbb{R}^3$ given by the solution space of the equation:

$$x_1 + 2x_2 + 3x_3 = 0$$

Find a basis for $W^\perp$, the orthogonal complement of W.

Answer available at the end of this section.

Exercise 4.14.E.3

Let W be the subspace of $\mathbb{R}^4$ spanned by:

$$\vec{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Find a basis for $W^\perp$, the orthogonal complement.

Exercise 4.14.E.4

Let W be the subspace of $\mathbb{R}^4$ spanned by the vectors:

$$\vec{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

Find a basis for the orthogonal complement $W^\perp$.

Answer available at the end of this section.

Exercise $4.14.E.5$

Consider the following scalar equation of a plane.

$$2x - 3y + z = 0$$

Find the point on the plane which is closest to $Y = (3, 4, 1)$ using the steps outlined in Example 4.10.5.

a. Find a basis X of the subspace W of $\mathbb{R}^3$ defined by the equation $2x - 3y + z = 0$.
b. Orthogonalize the basis X to get an orthogonal basis B of W.
c. Find the projection on W of the position vector of the point Y.

Exercise $4.14.E.6$

Consider the following scalar equation of a plane.

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

Find the point on the plane which is closest to $Y = (1, 2, 1)$.

Exercise $4.14.E.7$

Let $\vec{v}$ be a vector and let $\vec{n}$ be a normal vector for a plane through the origin. Find the equation of the line through the point determined by $\vec{v}$ which has direction vector $\vec{n}$. Show that it intersects the plane at the point determined by $\vec{v} - proj_{\vec{n}}\vec{v}$.

Hint available at the end of this section.

Exercise $4.14.E.8$

As shown in the above problem, one can find the closest point to $\vec{v}$ in a plane through the origin by finding the intersection of the line through $\vec{v}$ having direction vector equal to the normal vector to the plane with the plane. If the plane does not pass through the origin, this will still work to find the point on the plane closest to the point determined by $\vec{v}$. Here is a relation which defines a plane

$$2x + y + z = 11$$

and here is a point: $(1, 1, 2)$. Find the point on the plane which is closest to this point. Then determine the distance from the point to the plane by taking the distance between these two points.

Hint: available at the end of this section.

Exercise $4.14.E.9$

As shown in the above problem, one can find the closest point to~v in a plane through the origin by finding the intersection of the line through $\vec{v}$ having direction vector equal to the normal vector to the plane with the plane. If the plane does not pass through the origin, this will still work to find the point on the plane closest to the point determined by $\vec{v}$. Here is a relation which defines a plane

$$2x + y + z = 11$$

and here is a point: $(1, 1, 2)$. Find the point on the plane which is closest to this point. Then determine the distance from the point to the plane by taking the distance between these two points.

Hint: available at the end of this section.

Exercise $4.14.E.10$

Consider the plane

$$3x + 4y - 5z = 7$$

a. Find the point on this plane that is closest to the point $(2, -1, 3)$.
b. Compute the distance between $(2, -1, 3)$ and the plane.

Exercise $4.14.E.11$

Provide a proof for proposition 4.10.2: Let W be a subspace of $\mathbb{R}^n$. Then the orthogonal complement $W^\perp$ is also a subspace of $\mathbb{R}^n$.

Answers

Exercise $4.14.E.1$

$$\text{proj}_W \vec{v} = \begin{bmatrix} \frac{3}{2} \\ \frac{5}{2} \\ 1 \end{bmatrix}$$

Exercise $4.14.E.2$

$$W^\perp = \text{span}\left\{ \begin{bmatrix} 1 \\ -2 \\ -3 \end{bmatrix} \right\}$$

Exercise $4.14.E.4$

$$W^\perp = \text{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Exercise $4.14.E.7$

\[The line:$\vec{v} + t\vec{n}$. It is in the plane if $\vec{n} \bullet (\vec{v} + t\vec{n}) = 0$. Determine t. Then substitute in to the equation of the line.

Exercise $4.14.E.8$

Line: $(x, y, z) = (1, 1, 2) + t(2, 1, 1)$. Now require that it intersect the plane.

Exercise $4.14.E.9$

Line: $(x, y, z) = (1, 1, 2) + t(2, 1, 1)$. Now require that it intersect the plane.

This page titled 4.14.E: Exercise for Section 4.10 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

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

4.15: 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 4.15.1: 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:

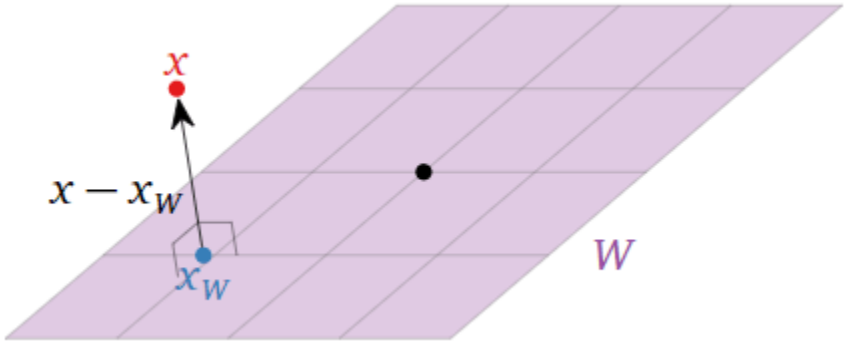


Figure 4.15.1

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 4.15.1: 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 available at the end of this section.

Definition 4.15.2: 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 4.15.1: 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 4.15.1: 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:

$$\begin{aligned} x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \implies x_W = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \quad x_{W^\perp} = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} \\ x = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \implies x_W = \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} \quad x_{W^\perp} = \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix}. \end{aligned} \tag{4.15.1}$$

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).

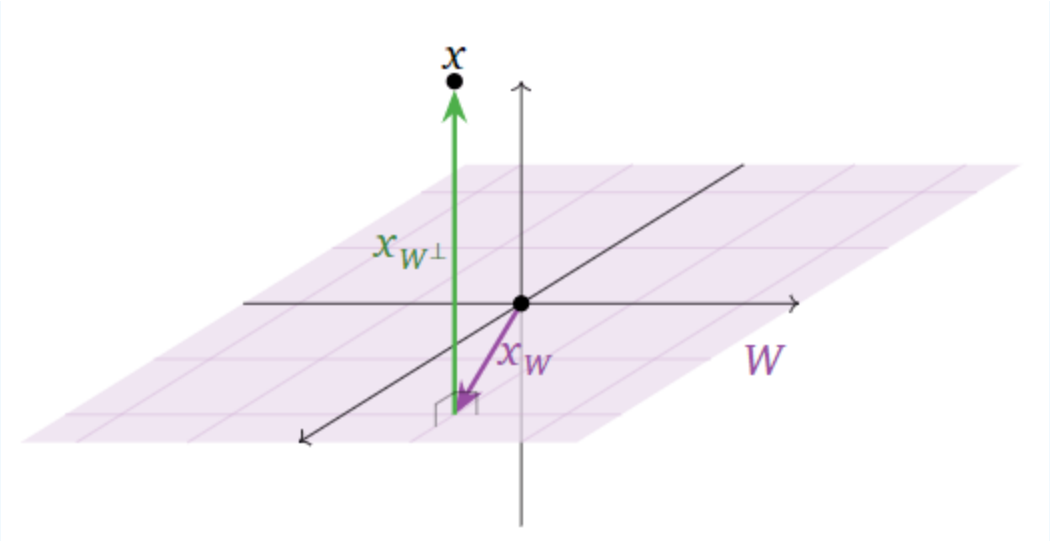

x
$x_{W^\perp}$
x_W
W

Figure 4.15.2

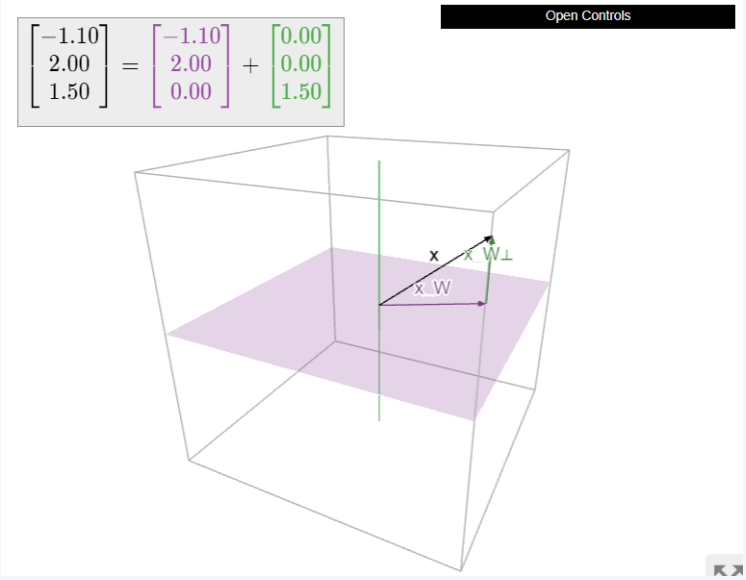

$\begin{bmatrix}-1.10\\2.00\\1.50\end{bmatrix} = \begin{bmatrix}-1.10\\2.00\\0.00\end{bmatrix} + \begin{bmatrix}0.00\\0.00\\1.50\end{bmatrix}$
Open Controls
x
x_W⊥
x_W

Figure 4.15.3: 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 4.15.2: 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 4.15.3: 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 4.15.4: Interactive: Orthogonal decomposition in $\mathbb{R}^2$

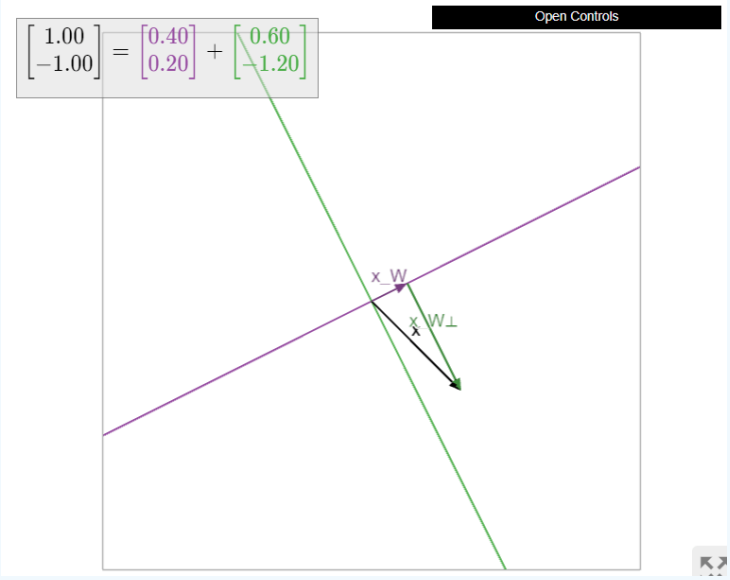


Figure 4.15.4: Orthogonal decomposition of a vector with respect to a line W in $\mathbb{R}^2$. Note that x_W is in W and $x_{W^\perp}$ is in the line perpendicular to W. Click and drag the head of the vector x to see how the orthogonal decomposition changes.

Example 4.15.5: Interactive: Orthogonal decomposition in $\mathbb{R}^3$

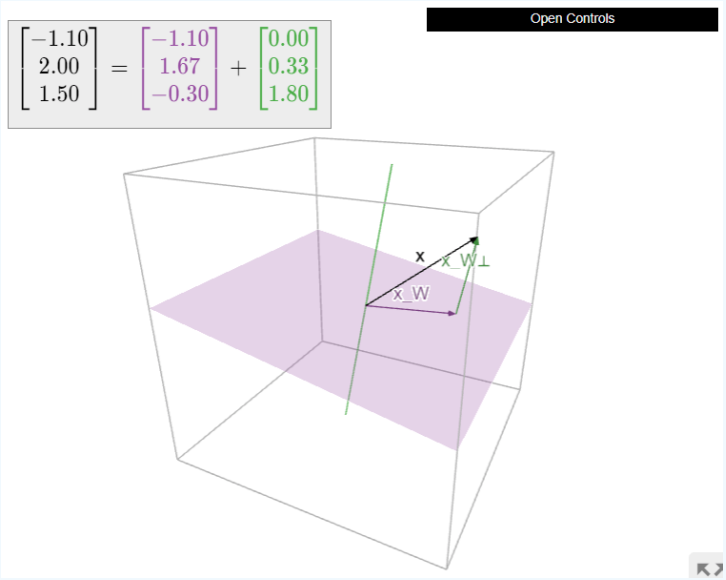


Figure 4.15.5: Orthogonal decomposition of a vector with respect to a plane W in $\mathbb{R}^3$. Note that x_W is in W and $x_{W^\perp}$ is in the line perpendicular to W. Click and drag the head of the vector x to see how the orthogonal decomposition changes.

Example 4.15.6: Interactive: Orthogonal decomposition in $\mathbb{R}^3$

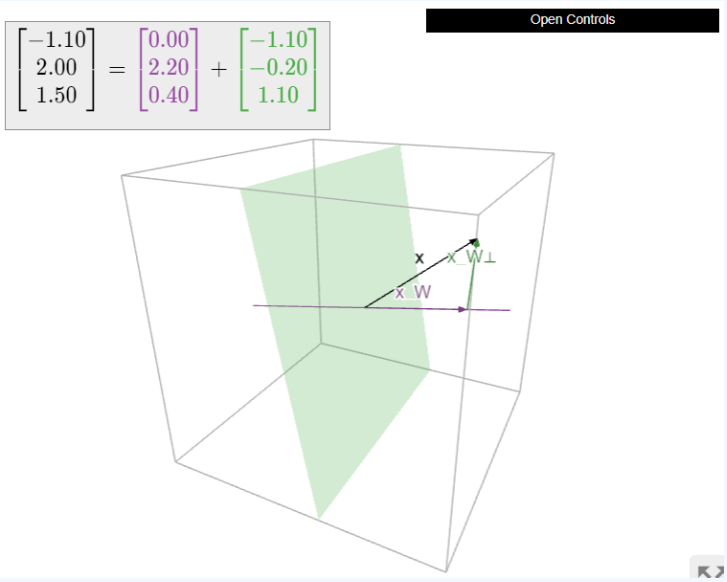


Figure 4.15.6: Orthogonal decomposition of a vector with respect to a line W in $\mathbb{R}^3$. Note that x_W is in W and $x_{W^\perp}$ is in the plane perpendicular to W. Click and drag the head of the vector x to see how the orthogonal decomposition changes.

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

Theorem 4.15.2

Let A be an $m \times n$ matrix, let $W = \mathrm{Col}(A)$, 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 available at the end of this section.

Example 4.15.7: Orthogonal Projection onto a Line

Let $L = \mathrm{Span}\{u\}$ be a line in $\mathbb{R}^n$ and let x be a vector in $\mathbb{R}^n$. By Theorem 4.15.2, 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 $x_L = uc = (u \cdot x)/(u \cdot u)\ u$.

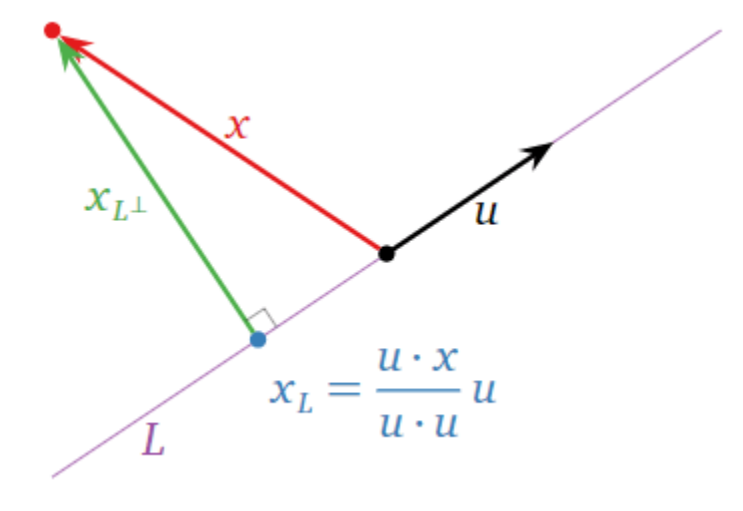


Figure 4.15.7

To reiterate:

Recipe: Orthogonal Projection onto a Line

If $L = \mathrm{Span}\{u\}$ 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 $L = \mathrm{Span}\{u\}$, 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.

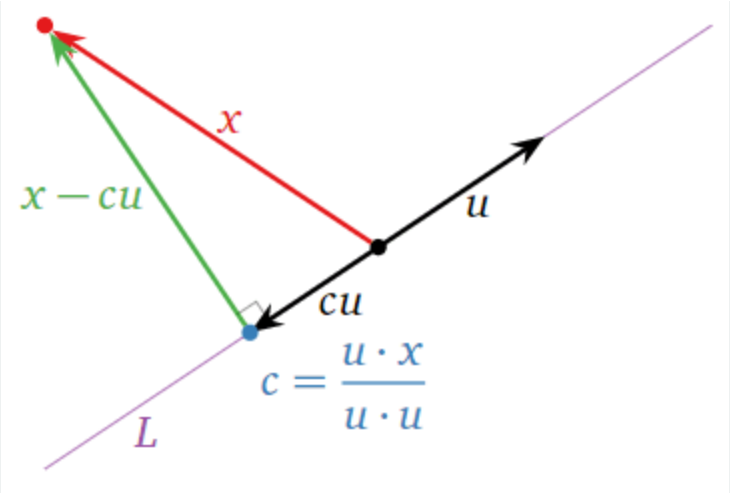


Figure 4.15.8

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 4.15.8: 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

$$\|x_{L^\perp}\| = \frac{1}{13}\sqrt{48^2 + 72^2} \approx 6.656.$$

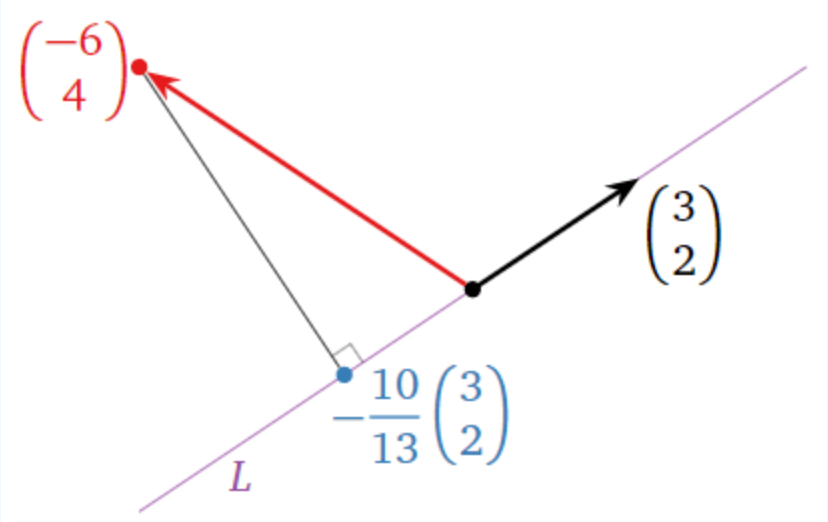


Figure 4.15.9

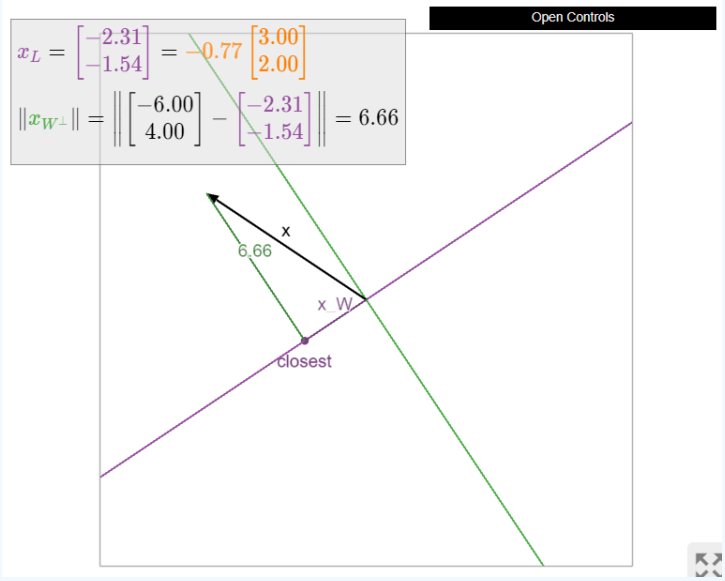


Figure 4.15.10: Distance from the line L.

Example 4.15.9: 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

$$x_L = \frac{x\cdot u}{u\cdot u}\,u = \frac{2+3-1}{1+1+1}\begin{pmatrix}-1\\1\\1\end{pmatrix} = \frac{4}{3}\begin{pmatrix}-1\\1\\1\end{pmatrix} \qquad x_{L^\perp} = x - x_L = \frac{1}{3}\begin{pmatrix}-2\\5\\-7\end{pmatrix}.$$

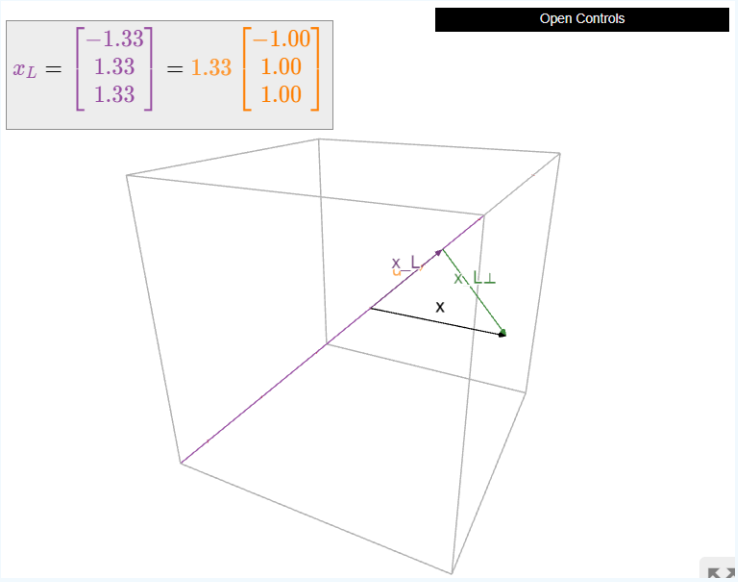


Figure 4.15.11: Orthogonal projection onto the line L.

When A is a matrix with more than one column, computing the orthogonal projection of x onto $W = \mathrm{Col}(A)$ means solving the matrix equation $A^TAc = A^Tx$. In other words, we can compute the closest vector by *solving a system of linear equations*. To be explicit, we state Theorem 4.15.2 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^TA and the vector A^Tx.
2. Form the augmented matrix for the matrix equation $A^TAc = A^Tx$ 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 4.15.10: Projection onto the xy-plane

Use Theorem 4.15.2 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 4.15.1.

Example 4.15.11: 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:

$$\begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & 3 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & -7/3 \\ 0 & 1 & 8/3 \end{pmatrix} \implies c = \frac{1}{3} \begin{pmatrix} -7 \\ 8 \end{pmatrix}.$$

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

$$\|x_{W^\perp}\| = \frac{1}{3}\sqrt{4+4+4} \approx 1.155.$$

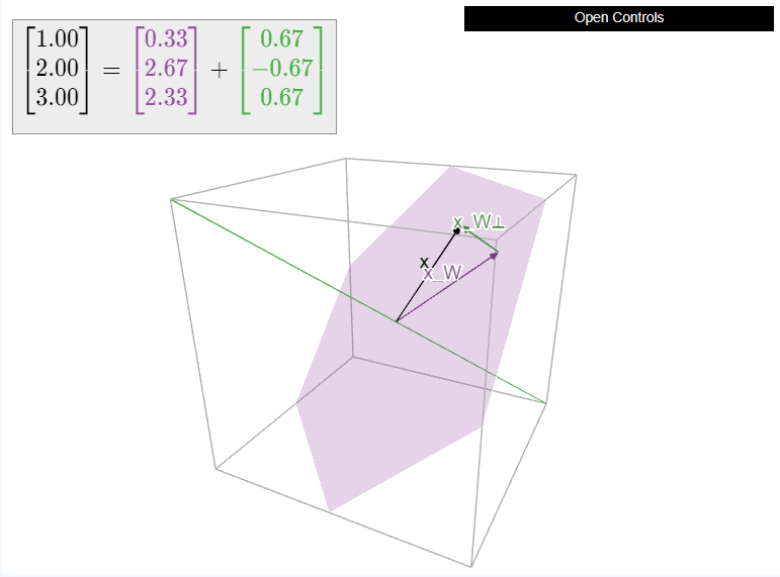


Figure 4.15.12: Orthogonal projection onto the plane W.

Example 4.15.12: 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 $\text{Col}(A) = \text{Nul}\begin{pmatrix} 1 & -2 & -1 \end{pmatrix} = W$.

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:

$$\begin{pmatrix} 5 & 2 & 3 \\ 2 & 2 & 2 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 1/3 \\ 0 & 1 & 2/3 \end{pmatrix}.$$

Hence we can take $c = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}$, so

$$x_W = Ac = \begin{pmatrix} 2 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 4 \\ 1 \\ 2 \end{pmatrix}.$$

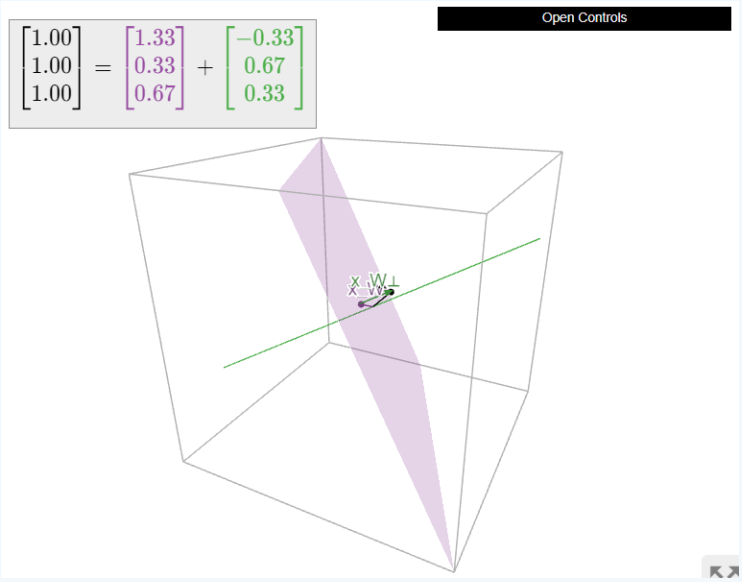


Figure 4.15.13: Orthogonal projection onto the plane W.

Method 2: In this case, it is easier to compute $x_{W^\perp}$. Indeed, since $W = \text{Nul}\begin{pmatrix} 1 & -2 & -1 \end{pmatrix}$, the orthogonal complement is the line

$$V = W^\perp = \text{Col}\begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}.$$

Using the formula for projection onto a line, Example 4.15.7, 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 4.15.13: 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:

$$\begin{pmatrix} 2 & 0 & 0 & -3 \\ 0 & 2 & 2 & -3 \\ 0 & 2 & 4 & 0 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 0 & -3/2 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 3/2 \end{pmatrix} \implies c = \frac{1}{2}\begin{pmatrix} -3 \\ -6 \\ 3 \end{pmatrix}.$$

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 4.15.1

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 available at the end of this section.

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 4.15.14: Computing a projection

Continuing with the above Example 4.15.11, 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 4.15.1, 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

$$\begin{aligned} x_W = A(A^TA)^{-1}A^Tx &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 0 \end{pmatrix}\frac{1}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= \frac{1}{3}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2x_1 + x_2 - x_3 \\ x_1 + 2x_2 + x_3 \\ -x_1 + x_2 + 2x_3 \end{pmatrix}. \end{aligned} \tag{4.15.2}$$

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 4.15.1: 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 available at the end of this section.

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 4.15.15: 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

$$\left.\begin{aligned} (e_1)_L &= \frac{u \cdot e_1}{u \cdot u} \quad u = \frac{3}{13}\begin{pmatrix} 3 \\ 2 \end{pmatrix} \\ (e_2)_L &= \frac{u \cdot e_2}{u \cdot u} \quad u = \frac{2}{13}\begin{pmatrix} 3 \\ 2 \end{pmatrix} \end{aligned}\right\} \quad \Longrightarrow \quad B = \frac{1}{13}\begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix}.$$

Example 4.15.16: 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

$$\left.\begin{aligned} (e_1)_L &= \frac{u \cdot e_1}{u \cdot u} \quad u = \frac{-1}{3}\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \\ (e_2)_L &= \frac{u \cdot e_2}{u \cdot u} \quad u = \frac{1}{3}\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \\ (e_3)_L &= \frac{u \cdot e_3}{u \cdot u} \quad u = \frac{1}{3}\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \end{aligned}\right\} \quad \Longrightarrow \quad B = \frac{1}{3}\begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix}.$$

Example 4.15.17: Matrix of a projection

Continuing with Example 4.15.11, let

$$W = \operatorname{Span}\left\{\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right\},$$

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:

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 1/3 \\ 0 & 1 & 1/3 \end{pmatrix} \implies (e_1)_W = A\begin{pmatrix} 1/3 \\ 1/3 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}.$$

We do the same for e_2:

$$\begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & -1/3 \\ 0 & 1 & 2/3 \end{pmatrix} \implies (e_1)_W = A\begin{pmatrix} -1/3 \\ 2/3 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

and for e_3:

$$\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 0 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & -2/3 \\ 0 & 1 & 1/3 \end{pmatrix} \implies (e_1)_W = A\begin{pmatrix} -2/3 \\ 1/3 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}.$$

It follows that

$$B = \frac{1}{3}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}.$$

In the previous Example 4.15.17, 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 4.15.1. In this case, we have already expressed T as a matrix transformation with matrix $A(A^TA)^{-1}A^T$. See this

Example 4.15.14.

Note 4.15.2

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 4.15.1, into properties of the associated standard matrix.

Proposition 4.15.2: 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 available at the end of this section.

We emphasize that the properties of projection matrices, Proposition 4.15.2, 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 4.15.17. 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 4.15.18

Continuing with above Example 4.15.17, 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

$$W = \mathrm{Span}\left\{\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right\}.$$

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

$$W^\perp = \mathrm{Span}\left\{\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\right\}.$$

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 4.15.18, 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 $\{v_1, v_2, \dots, v_m\}$ is a basis for W and $\{v_{m+1}, v_{m+2}, \dots, v_n\}$ 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 4.15.1.

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.

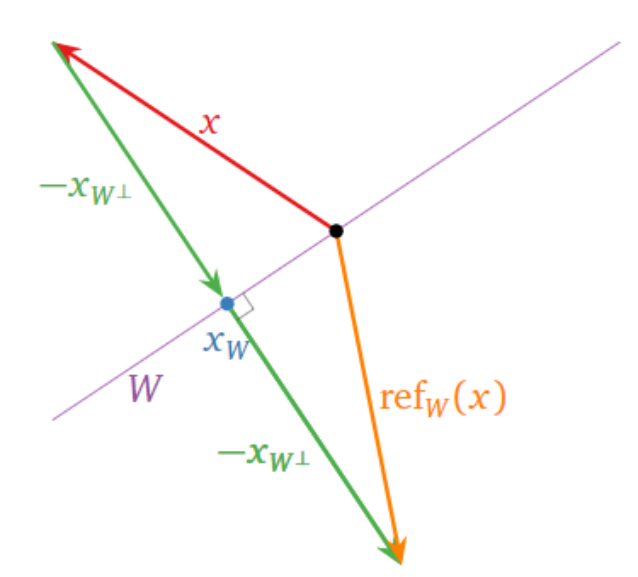


Figure 4.15.14

Since $x_{W^\perp} = x - x_W$, we also have

$$\mathrm{ref}_W(x) = x - 2(x - x_W) = 2x_W - x.$$

We leave it to the reader to check using the definition that:

1. $\mathrm{ref}_W \circ \mathrm{ref}_W = \mathrm{Id}_{\mathbb{R}^n}$.
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.

Answers

Theorem 4.15.1**: Orthogonal Decomposition**

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

$$x = (c_1v_1 + \cdots + c_mv_m) + (c_{m+1}v_{m+1} + \cdots + c_nv_n) = x_W + x_{W^\perp},$$

where $x_W = c_1v_1 + \cdots + c_mv_m$ and $x_{W^\perp} = c_{m+1}v_{m+1} + \cdots + c_nv_n$. 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.

Theorem4.15.2

Let $x = x_W + x_{W^\perp}$ be the orthogonal decomposition with respect to W. By definition x_W lies in $W = \mathrm{Col}(A)$ 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 $\mathrm{Nul}(A^T)$ by Recipe: Shortcuts for Computing Orthogonal Complements in Section 6.2. We thus have

$$0 = A^T(x - Ac) = A^Tx - A^TAc$$

and so

$$A^TAc = A^Tx.$$

This exactly means that $A^TAc = A^Tx$ is consistent. If c is any solution to $A^TAc = A^Tx$ then by reversing the above logic, we conclude that $x_W = Ac$.

Corollary 4.15.1

We will show that $\mathrm{Nul}(A^TA) = \{0\}$, 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 4.15.2. But $0_W = 0$ (the orthogonal decomposition of the zero vector is just $0 = 0 + 0$), so $Ac = 0$, and therefore c is in $\mathrm{Nul}(A)$. Since the columns of A are linearly independent, we have $c = 0$, so $\mathrm{Nul}(A^TA) = 0$, 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 $x_W = Ac = A(A^TA)^{-1}A^Tx$.

Proposition $4.15.1$**: Properties of Orthogonal Projections**

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 $4.15.2$.
3. See Example $4.15.3$.
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.

Proposition $4.15.2$**: Properties of Projection Matrices**

The first four assertions are translations of properties 5, 3, 4, and 2 from Proposition $4.15.2$, 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 $4.15.1$ that there is a basis of $\mathbb{R}^n$ of the form $\{v_1, \dots, v_m, v_{m+1}, \dots, v_n\}$, where $\{v_1, \dots, v_m\}$ is a basis for W and $\{v_{m+1}, \dots, v_n\}$ 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 $1, \dots, 1, 0, \dots, 0$ (m ones and $n - m$ zeros). Now we use Theorem 5.4.1 in Section 5.4.

This page titled 4.15: Orthogonal Projection is shared under a GNU FDL license and was authored, remixed, and/or curated by Dan Margalit and Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- 6.3: Orthogonal Projection by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

4.16: Least Squares Approximation

Learning Objectives

- Find the least squares solution for an inconsistent system.
- Apply the least squares approximate to find the regression line given a collection of points.

It should not be surprising to hear that many problems do not have a perfect solution, and in these cases the objective is always to try to do the best possible. For example what does one do if there are no solutions to a system of linear equations $A\vec{x} = \vec{b}$? It turns out that what we do is find $\vec{x}$ such that $A\vec{x}$ is as close to $\vec{b}$ as possible. A very important technique that follows from orthogonal projections is that of the least square approximation, and allows us to do exactly that.

We begin with a lemma.

Recall that we can form the image of an $m \times n$ matrix A by $\text{im}\,(A) = \{A\vec{x} : \vec{x}^{\,\text{int}}\mathbb{R}^n\}$. Rephrasing Theorem $4.16.4$ using the subspace $W = \text{im}\,(A)$ gives the equivalence of an orthogonality condition with a minimization condition. The following picture illustrates this orthogonality condition and geometric meaning of this theorem.

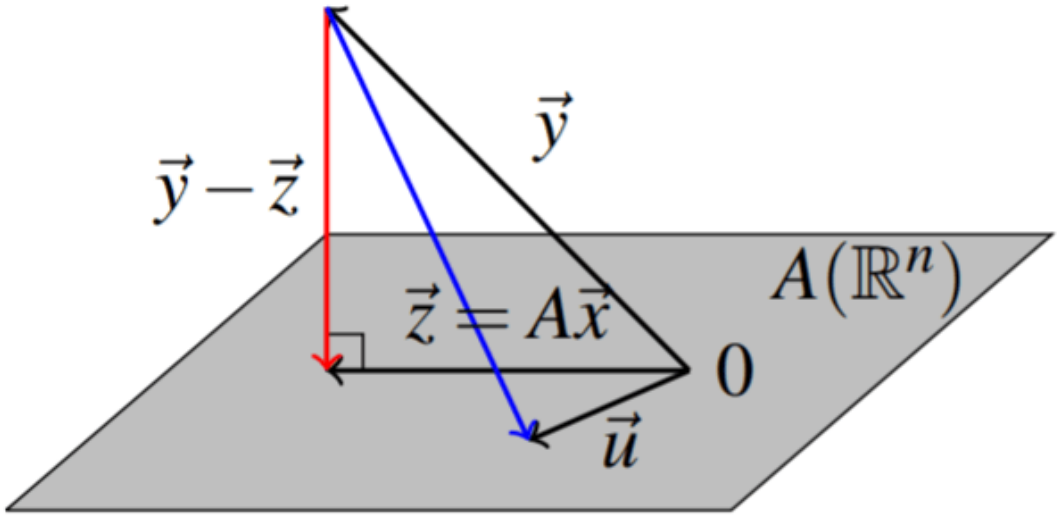


Figure $4.16.2$ A geometric representation of least squares regression. A gray plane represents the column space $A(\mathbb{R}^n)$ of matrix A, containing the origin 0. A black vector $\vec{y}$ extends from the origin to a point above the plane. The projection $\vec{z} = A\vec{x}$ is shown as a blue vector from the origin to the plane. A red vector $\vec{y} - \vec{z}$ is perpendicular to the plane, representing the residual. Another black vector $\vec{u}$ lies in the plane, completing the triangle. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

Theorem $4.16.1$: Existence of Minimizers

Let $\vec{y}^{\,\text{int}}\mathbb{R}^m$ and let A be an $m \times n$ matrix.

Choose $\vec{z}^{\,\text{int}}W = \text{im}\,(A)$ given by $\vec{z} = \text{proj}_W(\vec{y})$, and let $\vec{x}^{\,\text{int}}\mathbb{R}^n$ such that $\vec{z} = A\vec{x}$.

Then

1. $\vec{y} - A\vec{x}^{\,\text{int}}W^{\perp}$
2. $\|\vec{y} - A\vec{x}\| < \|\vec{y} - \vec{u}\|$ for all $\vec{u} \neq \vec{z}^{\,\text{int}}W$

We note a simple but useful observation.

Lemma 4.16.1: Transpose and Dot Product

Let A be an $m \times n$ matrix. Then

$$A\vec{x} \cdot \vec{y} = \vec{x} \cdot A^T\vec{y}$$

Proof available at the end of this section.

The next corollary gives the technique of least squares.

Corollary 4.16.1: Least Squares and Normal Equation

A specific value of $\vec{x}$ which solves the problem of Theorem 4.16.5 is obtained by solving the equation

$$A^TA\vec{x} = A^T\vec{y}$$

Furthermore, there always exists a solution to this system of equations.

Proof available at the end of this section.

Note that $\vec{x}$ might not be unique but $A\vec{x}$, the closest point of $A(\mathbb{R}^n)$ to $\vec{y}$ is unique as was shown in the above argument.

Consider the following example.

Example 4.16.1: Least Squares Solution to a System

Find a least squares solution to the system

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

Solution

First, consider whether there exists a real solution. To do so, set up the augmnented matrix given by

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

The reduced row-echelon form of this augmented matrix is

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

It follows that there is no real solution to this system. Therefore we wish to find the least squares solution. The normal equations are

$$A^TA\vec{x} = A^T\vec{y}$$

$$\begin{bmatrix} 2 & -1 & 4 \\ 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 & -1 & 4 \\ 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \tag{4.16.1}$$

and so we need to solve the system

$$\begin{bmatrix} 21 & 19 \\ 19 & 35 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 10 \end{bmatrix}$$

This is a familiar exercise and the solution is

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{5}{34} \\ \frac{7}{34} \end{bmatrix}$$

Consider another example.

Example 4.16.2: Least Squares Solution to a System

Find a least squares solution to the system

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

Solution

First, consider whether there exists a real solution. To do so, set up the augmnented matrix given by

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

The reduced row-echelon form of this augmented matrix is

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

It follows that the system has a solution given by $x = y = 1$. However we can also use the normal equations and find the least squares solution.

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

Then

$$\begin{bmatrix} 21 & 19 \\ 19 & 35 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 40 \\ 54 \end{bmatrix}$$

The least squares solution is

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

which is the same as the solution found above.

An important application of Corollary 4.16.1 is the problem of finding the least squares regression line in statistics. Suppose you are given points in the xy plane

$$\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$$

and you would like to find constants m and b such that the line $\vec{y} = m\vec{x} + b$ goes through all these points. Of course this will be impossible in general. Therefore, we try to find m, b such that the line will be as close as possible. The desired system is

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix}$$

which is of the form $\vec{y} = A\vec{x}$. It is desired to choose m and b to make

$$\left\| A \begin{bmatrix} m \\ b \end{bmatrix} - \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \right\|^2$$

as small as possible. According to Theorem $4.16.5$ and Corollary $4.16.1$, the best values for m and b occur as the solution to

$$A^T A \begin{bmatrix} m \\ b \end{bmatrix} = A^T \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \text{where } A = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}$$

Thus, computing $A^T A$,

$$\begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{bmatrix}$$

Solving this system of equations for m and b (using Cramer's rule for example) yields:

$$m = \frac{-\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right) + \left(\sum_{i=1}^n x_i y_i\right) n}{\left(\sum_{i=1}^n x_i^2\right) n - \left(\sum_{i=1}^n x_i\right)^2}$$

and

$$b = \frac{-\left(\sum_{i=1}^n x_i\right) \sum_{i=1}^n x_i y_i + \left(\sum_{i=1}^n y_i\right) \sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i^2\right) n - \left(\sum_{i=1}^n x_i\right)^2}.$$

Consider the following example.

Example $4.16.3$: Least Squares Regression

Find the least squares regression line $y = mx + b$ for the following set of data points:

$$\{(0,1),(1,2),(2,2),(3,4),(4,5)\}$$

Solution

In this case we have $n = 5$ data points and we obtain:

$$\sum_{i=1}^5 x_i = 10 \qquad \sum_{i=1}^5 y_i = 14$$

$$\sum_{i=1}^5 x_i y_i = 38 \quad \sum_{i=1}^5 x_i^2 = 30$$

and hence

$$\begin{aligned} m &= \frac{-10 * 14 + 5 * 38}{5 * 30 - 10^2} = 1.00 \\ b &= \frac{-10 * 38 + 14 * 30}{5 * 30 - 10^2} = 0.80 \end{aligned} \tag{4.16.2}$$

The least squares regression line for the set of data points is:

$$y = x + .8$$

One could use this line to approximate other values for the data. For example for $x = 6$ one could use $y(6) = 6 + .8 = 6.8$ as an approximate value for the data.

The following diagram shows the data points and the corresponding regression line.

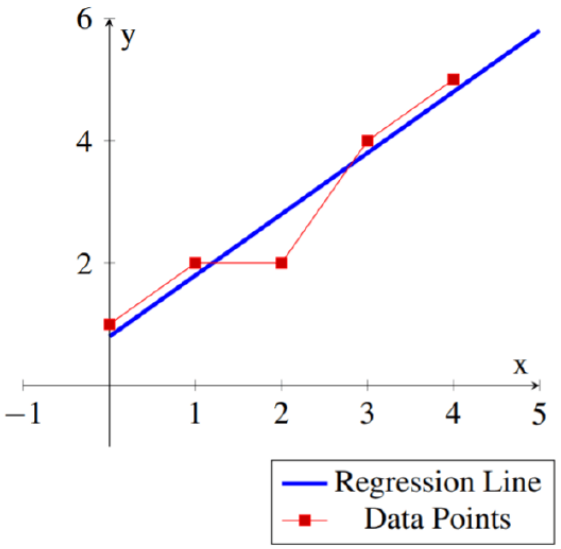


Figure 4.16.3 The diagram shows the data points $(0,1),(1,2),(2,2),(3,4),(4,5)$ and the corresponding regression line $y=x+.8$. (CC BY-NC-SA 4.0; Kuttler via A First Course in LINEAR ALGEBRA)

One could clearly do a least squares fit for curves of the form $y=ax^2+bx+c$ in the same way. In this case you want to solve as well as possible for a,b, and c the system

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

and one would use the same technique as above. Many other similar problems are important, including many in higher dimensions and they are all solved the same way.

Answers

Lemma 4.16.1: Transpose and Dot Product

This follows from the definitions:

$$A\vec{x}\cdot\vec{y}=\sum_{i,j}a_{ij}x_jy_i=\sum_{i,j}x_ja_{ji}y_i=\vec{x}\cdot A^T\vec{y}$$

Corollary 4.16.1: Least Squares and Normal Equation

For $\vec{x}$ the minimizer of Theorem 4.16.5, $(\vec{y}-A\vec{x})\cdot A\vec{u}=0$ for all $\vec{u}^{\text{int}}\mathbb{R}^n$ and from Lemma 4.16.1, this is the same as saying

$$A^T(\vec{y}-A\vec{x})\cdot\vec{u}=0$$

for all $u^{\text{int}}\mathbb{R}^n$. This implies

$$A^T\vec{y}-A^TA\vec{x}=\vec{0}.$$

Therefore, there is a solution to the equation of this corollary, and it solves the minimization problem of Theorem 4.16.5.

This page titled 4.16: Least Squares Approximation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

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

4.16.E: Exercises for Section 4.11

Exercise 4.16.E.1

Find the least squares solution to the following system.

$$\begin{aligned} x+2y &= 1 \\ 2x+3y &= 2 \\ 3x+5y &= 4 \end{aligned} \tag{4.16.E.1}$$

Answer available at the end of this section.

Exercise 4.16.E.2

You are doing experiments and have obtained the ordered pairs,

$$(0,1),(1,2),(2,3.5),(3,4)$$

Find m and b such that $y=mx+b$ approximates these four points as well as possible.

Exercise 4.16.E.3

Given the ordered pairs $(2,1)$, $(1,2)$, $(0,1)$,$(-1,2)$, and $(-2,1)$, find m and b such that $y=mx+b$ approximates these five points as well as possible.

Answer available at the end of this section.

Exercise 4.16.E.4

Suppose you have several ordered triples, (x_i,y_i,z_i). Describe how to find a polynomial such as

$$z=a+bx+cy+dxy+ex^2+fy^2$$

giving the best fit to the given ordered triples.

Exercise 4.16.E.5

Find the least squares solution for the system

$$\begin{aligned} x+y+z &= 2 \\ 2x+3y+4z &= 5 \\ 3x+5y+7z &= 8 \\ 4x+7y+10z &= 11 \end{aligned} \tag{4.16.E.2}$$

Exercise 4.16.E.6

In general, you have a point (x_0,y_0,z_0) and a scalar equation for a plane $ax+by+cz=d$ where $a^2+b^2+c^2>0$. Determine a formula for the closest point on the plane to the given point. Then use this point to get a formula for the distance from the given point to the plane.

Answer available at the end of this section.

Answers

Exercise $4.16.E.1$

$$\begin{aligned}\begin{bmatrix}1&2\\2&3\\3&5\end{bmatrix}^T\begin{bmatrix}1&2\\2&3\\3&5\end{bmatrix}&=\begin{bmatrix}14&23\\23&38\end{bmatrix}\begin{bmatrix}14&23\\23&38\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}\\&=\begin{bmatrix}1&2\\2&3\\3&5\end{bmatrix}^T\begin{bmatrix}1\\2\\4\end{bmatrix}=\begin{bmatrix}17\\28\end{bmatrix}\end{aligned}\tag{4.16.E.3}$$

$$\begin{aligned}\begin{bmatrix}14&23\\23&38\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}&=\begin{bmatrix}17\\28\end{bmatrix}\\\begin{bmatrix}14&23\\23&38\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}&=\begin{bmatrix}17\\28\end{bmatrix},\end{aligned}\tag{4.16.E.4}$$

Solution is: $\begin{bmatrix}\frac{2}{3}\\\frac{1}{3}\end{bmatrix}$

Exercise $4.16.E.3$

$y=1.4$

Exercise $4.16.E.6$

Find the line perpendicular to the plane which goes through the given point: $(x,y,z)=(x_0,y_0,z_0)+t(a,b,c)$. Now require that this point satisfy the equation for the plane to determine t.

This page titled 4.16.E: Exercises for Section 4.11 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Doli Bambhania, Fatemeh Yarahmadi, and Bill Wilson.

4.17: Vector Spaces and Subspace

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 4.17.1: 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 4.17.2: 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

$$\left(\vec{v} + \vec{w}\right) + \vec{z} = \vec{v} + \left(\vec{w} + \vec{z}\right)$$

- The Existence of an Additive Identity

$$\vec{v} + \vec{0} = \vec{v}$$

- The Existence of an Additive Inverse

$$\vec{v} + \left(-\vec{v}\right) = \vec{0}$$

vectorspaceaxiomsaddition

Definition 4.17.3: Axioms of Scalar Multiplication

Let $a, b^{\text{int}}\mathbb{R}$ 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

$$\text{If } a \text{ is a real number, and } \vec{v} \text{ is in } V, \text{ then } a\vec{v} \text{ is in } V.$$

- $$a\left(\vec{v}+\vec{w}\right)=a\vec{v}+a\vec{w}$$
- $$\left(a+b\right)\vec{v}=a\vec{v}+b\vec{v}$$
- $$a\left(b\vec{v}\right)=\left(ab\right)\vec{v}$$
- $$1\vec{v}=\vec{v}$$

Consider the following example, in which we prove that $\mathbb{R}^n$ is in fact a vector space.

Example 4.17.1: $\mathbb{R}^n$

$\mathbb{R}^n$, 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:

$$\begin{aligned} \vec{x}+\vec{y} &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ &= \begin{bmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{bmatrix} \\ &= \begin{bmatrix} y_1+x_1 \\ y_2+x_2 \\ \vdots \\ y_n+x_n \end{bmatrix} \\ &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \vec{y}+\vec{x} \end{aligned} \tag{4.17.1}$$

  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.

$$
\begin{aligned}
(\vec{x}+\vec{y})+\vec{z} &= \left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}\right) + \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \\
&= \begin{bmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \\
&= \begin{bmatrix} (x_1+y_1)+z_1 \\ (x_2+y_2)+z_2 \\ \vdots \\ (x_n+y_n)+z_n \end{bmatrix} \\
&= \begin{bmatrix} x_1+(y_1+z_1) \\ x_2+(y_2+z_2) \\ \vdots \\ x_n+(y_n+z_n) \end{bmatrix} \\
&= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1+z_1 \\ y_2+z_2 \\ \vdots \\ y_n+z_n \end{bmatrix} \\
&= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}\right) \\
&= \vec{x}+(\vec{y}+\vec{z})
\end{aligned}
\tag{4.17.2}
$$

Hence addition of vectors is associative.

- Next, we show the existence of an additive identity. Let $\vec{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

$$
\begin{aligned}
\vec{x}+\vec{0} &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} x_1+0 \\ x_2+0 \\ \vdots \\ x_n+0 \end{bmatrix} \\
&= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= \vec{x}
\end{aligned}
\tag{4.17.3}
$$

Hence the zero vector $\vec{0}$ is an additive identity.

- Next, we prove the existence of an additive inverse. Let $-\vec{x} = \begin{pmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{pmatrix}$.

$$\begin{aligned}\vec{x}+(-\vec{x}) &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 - x_1 \\ x_2 - x_2 \\ \vdots \\ x_n - x_n \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= \vec{0}\end{aligned} \tag{4.17.4}$$

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

$$\begin{aligned}a(\vec{x}+\vec{y}) &= a\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \\ \vdots x_n \end{bmatrix} \right) \\ &= a \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \\ &= \begin{bmatrix} a(x_1 + y_1) \\ a(x_2 + y_2) \\ \vdots \\ a(x_n + y_n) \end{bmatrix} \\ &= \begin{bmatrix} ax_1 + ay_1 \\ ax_2 + ay_2 \\ \vdots \\ ax_n + ay_n \end{bmatrix} \\ &= \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix} + \begin{bmatrix} ay_1 \\ ay_2 \\ \vdots \\ ay_n \end{bmatrix} \\ &= a\vec{x} + a\vec{y}\end{aligned} \tag{4.17.5}$$

- Next, we wish to show that $(a+b)\vec{x} = a\vec{x} + b\vec{x}$.

Illegal pream-token (x)

- We wish to show that $a(b\vec{x}) = (ab)\vec{x}$.

$$\begin{aligned} a(b\vec{x}) &= a\left(b\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) \\ &= a\left(\begin{bmatrix} bx_1 \\ bx_2 \\ \vdots \\ bx_n \end{bmatrix}\right) \\ &= \begin{bmatrix} a(bx_1) \\ a(bx_2) \\ \vdots \\ a(bx_n) \end{bmatrix} \\ &= \begin{bmatrix} (ab)x_1 \\ (ab)x_2 \\ \vdots \\ (ab)x_n \end{bmatrix} \\ &= (ab)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= (ab)\vec{x} \end{aligned} \tag{4.17.6}$$

- Finally, we need to show that $1\vec{x} = \vec{x}$.

$$\begin{aligned} 1\vec{x} &= 1\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \begin{bmatrix} 1x_1 \\ 1x_2 \\ \vdots \\ 1x_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \vec{x} \end{aligned} \tag{4.17.7}$$

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 4.17.2: 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

$$\mathbb{P}_2 = \left\{a_2x^2 + a_1x + a_0 | a_i^{\text{int}}\mathbb{R} \text{ for all } i\right\}$$

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.

$$\begin{aligned} p(x) + q(x) &= p_2x^2 + p_1x + p_0 + q_2x^2 + q_1x + q_0 \\ &= (p_2 + q_2)x^2 + (p_1 + q_1)x + (p_0 + q_0) \end{aligned} \tag{4.17.8}$$

  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 $p(x) + q(x) = q(x) + p(x)$.

$$\begin{aligned} p(x) + q(x) &= p_2x^2 + p_1x + p_0 + q_2x^2 + q_1x + q_0 \\ &= (p_2 + q_2)x^2 + (p_1 + q_1)x + (p_0 + q_0) \\ &= (q_2 + p_2)x^2 + (q_1 + p_1)x + (q_0 + p_0) \\ &= q_2x^2 + q_1x + q_0 + p_2x^2 + p_1x + p_0 \\ &= q(x) + p(x) \end{aligned} \tag{4.17.9}$$

- Next, we need to show that addition is associative. That is, that $(p(x) + q(x)) + r(x) = p(x) + (q(x) + r(x))$.

$$\begin{aligned} (p(x) + q(x)) + r(x) &= \left(p_2x^2 + p_1x + p_0 + q_2x^2 + q_1x + q_0\right) + r_2x^2 + r_1x + r_0 \\ &= (p_2 + q_2)x^2 + (p_1 + q_1)x + (p_0 + q_0) + r_2x^2 + r_1x + r_0 \\ &= (p_2 + q_2 + r_2)x^2 + (p_1 + q_1 + r_1)x + (p_0 + q_0 + r_0) \\ &= p_2x^2 + p_1x + p_0 + (q_2 + r_2)x^2 + (q_1 + r_1)x + (q_0 + r_0) \\ &= p_2x^2 + p_1x + p_0 + \left(q_2x^2 + q_1x + q_0 + r_2x^2 + r_1x + r_0\right) \\ &= p(x) + (q(x) + r(x)) \end{aligned} \tag{4.17.10}$$

- Next, we must prove that there exists an additive identity. Let $0(x) = 0x^2 + 0x + 0$.

$$\begin{aligned} p(x) + 0(x) &= p_2x^2 + p_1x + p_0 + 0x^2 + 0x + 0 \\ &= (p_2 + 0)x^2 + (p_1 + 0)x + (p_0 + 0) \\ &= p_2x^2 + p_1x + p_0 \\ &= p(x) \end{aligned} \tag{4.17.11}$$

  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:

$$\begin{aligned} p(x) + (-p(x)) &= p_2x^2 + p_1x + p_0 + \left(-p_2x^2 - p_1x - p_0\right) \\ &= (p_2 - p_2)x^2 + (p_1 - p_1)x + (p_0 - p_0) \\ &= 0x^2 + 0x + 0 \\ &= 0(x) \end{aligned} \tag{4.17.12}$$

  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 $a(p(x) + q(x)) = ap(x) + aq(x)$.

$$\begin{aligned} a(p(x) + q(x)) &= a\left(p_2x^2 + p_1x + p_0 + q_2x^2 + q_1x + q_0\right) \\ &= a\left((p_2 + q_2)x^2 + (p_1 + q_1)x + (p_0 + q_0)\right) \\ &= a(p_2 + q_2)x^2 + a(p_1 + q_1)x + a(p_0 + q_0) \\ &= (ap_2 + aq_2)x^2 + (ap_1 + aq_1)x + (ap_0 + aq_0) \\ &= ap_2x^2 + ap_1x + ap_0 + aq_2x^2 + aq_1x + aq_0 \\ &= ap(x) + aq(x) \end{aligned} \tag{4.17.13}$$

- Next we show that $(a + b)p(x) = ap(x) + bp(x)$.

$$\begin{aligned}(a+b)p(x) &= (a+b)(p_2x^2+p_1x+p_0)\\ &= (a+b)p_2x^2+(a+b)p_1x+(a+b)p_0\\ &= ap_2x^2+ap_1x+ap_0+bp_2x^2+bp_1x+bp_0\\ &= ap(x)+bp(x)\end{aligned} \tag{4.17.14}$$

- The next axiom which needs to be verified is $a(bp(x)) = (ab)p(x)$.

$$\begin{aligned}a(bp(x)) &= a\left(b\left(p_2x^2+p_1x+p_0\right)\right)\\ &= a\left(bp_2x^2+bp_1x+bp_0\right)\\ &= abp_2x^2+abp_1x+abp_0\\ &= (ab)(p_2x^2+p_1x+p_0)\\ &= (ab)p(x)\end{aligned} \tag{4.17.15}$$

- Finally, we show that $1p(x) = p(x)$.

$$\begin{aligned}1p(x) &= 1\left(p_2x^2+p_1x+p_0\right)\\ &= 1p_2x^2+1p_1x+1p_0\\ &= p_2x^2+p_1x+p_0\\ &= p(x)\end{aligned} \tag{4.17.16}$$

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 4.17.3: 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.

$$\begin{aligned}A+B &= \begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\end{bmatrix}+\begin{bmatrix}b_{11} & b_{12} & b_{13}\\ b_{21} & b_{22} & b_{23}\end{bmatrix}\\ &= \begin{bmatrix}a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13}\\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23}\end{bmatrix}\end{aligned} \tag{4.17.17}$$

  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.

$$\begin{aligned}cA &= c\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\end{bmatrix}\\ &= \begin{bmatrix}ca_{11} & ca_{12} & ca_{13}\\ ca_{21} & ca_{22} & ca_{23}\end{bmatrix}\end{aligned} \tag{4.17.18}$$

  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 4.17.4: 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

$$\begin{aligned} A + B &= A \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ B + A &= B \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{aligned} \tag{4.17.19}$$

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 4.17.5: 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

$$\begin{aligned} (f + g)(x) &= f(x) + g(x) \\ (af)(x) &= a(f(x)) \end{aligned} \tag{4.17.20}$$

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 4.17.1: 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 available at the end of this section.

An important use of the additive inverse is the following theorem.

Theorem 4.17.2

Let V be a vector space. Then $\vec{v}+\vec{w}=\vec{v}+\vec{z}$ implies that $\vec{w}=\vec{z}$ for all $\vec{v},\vec{w},\vec{z}\,^{\text{int}}V$

Proof available at the end of this section.

Answers

Theorem 4.17.1**: Uniqueness**

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

   $$\begin{aligned} 0\vec{x} + (-0\vec{x}) &= 0\vec{x} + 0\vec{x} + (-\vec{x}) \\ \vec{0} + 0 &= 0\vec{x} + 0 \\ \vec{0} &= 0\vec{x} \end{aligned} \tag{4.17.21}$$

   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:

   $$\begin{aligned} (-1)\,\vec{x} + \vec{x} &= (-1)\,\vec{x} + 1\vec{x} \\ &= (-1 + 1)\,\vec{x} \\ &= 0\vec{x} \\ &= \vec{0} \end{aligned} \tag{4.17.22}$$

   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 $(-1)\,\vec{x} = -\vec{x}$.

Theorem 4.17.2

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 4.17: Vector Spaces and Subspace is shared under a license and was authored, remixed, and/or curated by LibreTexts.

- 9.1: Algebraic Considerations by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.

4.18: Amira Vector Space Axioms

Learning Objectives

By the end of this section, students will be able to:

- Identify the zero element and the additive inverse in a given set when nonstandard operations are defined.
- Verify closure under addition and scalar multiplication using arbitrary elements.
- Explain how different operation rules affect the structure of a set and determine whether it forms a vector space.

Why we Should Study Vector Spaces?

Previously, we worked with matrices and solved systems of linear equations, finding solutions and solution sets in $\mathbb{R}^n$. In those problems, we were allowed to add real numbers and multiply them by scalars. We also explored matrix addition and scalar multiplication, noting that these operations are only defined for matrices of the same size.

We discovered that **square matrices** play a special role in solving systems of equations, they follow consistent algebraic rules that allow us to perform operations like finding determinants and inverses.

What makes the set of all points in $\mathbb{R}^n$ and the set of all square matrices of a given size so powerful is that **addition and scalar multiplication always produce another element of the same set.** This property, staying "inside the set" after performing these operations, is one of the key ideas behind *vector spaces*.

The diagram below shows the **pitch control system** of the Space Shuttle, adapted from Figure 3 in the NASA Technical Memorandum *Space Shuttle Digital Flight Control System* (1976). It illustrates how feedback signals are combined to maintain the desired pitch angle during ascent. The diagram includes several summing junctions, represented by the symbol $\otimes$, where signals such as the commanded pitch and the measured pitch are added or subtracted. Each junction $\otimes$demonstrates the **closure property** of a vector space, meaning that when two valid signals are added, the result is another valid signal of the same type.

The system also includes blocks that multiply signals by constants, showing **scalar multiplication**. Together, the junctions $\otimes$ and the scaling blocks illustrate how the collection of all control signals, including commands, sensor readings, and feedback corrections, forms a **vector space of continuous functions**. This example connects the abstract definition of a vector space to a real engineering system that depends on addition and scaling of signals to maintain stability and performance.

Definition 4.18.1: Vector Space

A **Vector Space** is a set V of elements called **vectors**, having operations of addition and scalar multiplication defined on it that satisfy the following conditions (u, v, and w are arbitrary elements of V, and c and d are scalars.)

Closure Axioms

1. The sum $u+v$ exists and is an element of V. (V is closed under addition.)
2. cu is an element of V. (V is closed under multiplication.)

Addition Axioms

1. $u+v=v+u$ (commutative property)
2. $u+(v+w)=(u+v)+w$ (associative property)
3. There exists an element of V, called a **zero vector**, denoted $\mathbb{0}$, such that $u+0=u$

4. For every element u of V, there exists an element called a **negative** of u, denoted $-u$, such that $u + (-u) = 0$.

Scalar Multiplication Axioms

1. $c(u + v) = cu + cv$
2. $(c + d)u = cu + du$
3. $c(du) = (cd)u$
4. $1u = u$

In the 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.

Points to Remember When Checking a Vector Space

- The definition requires that (u, v, and w be **arbitrary (random) elements** of the set implies that
  - We cannot use specific examples or numbers to verify the axioms.
  - The properties must hold for **all** elements in the set.
- When testing if a set is a vector space, **always start** by checking the **closure properties**:
  1. **Closure under addition**
  2. **Closure under scalar multiplication**
  - These two checks often give a quick answer about whether the set can be a vector space.
- The operations of addition and scalar multiplication might **not be the usual ones**.
  - Because of this, the **zero element** may also be **different from the usual zero**.
  - You may need to **solve for the additive inverse** based on the operations defined in the problem.

Below are **examples and non-examples of vector spaces**. It is crucial to focus on how **randomness** (using arbitrary elements) is applied when showing whether a set is a vector space. In every case, we must test the properties using **general elements**, not specific numbers, to verify that the axioms hold for all possible cases.

Example 4.18.1: $\mathbb{R}^n$

$\mathbb{R}^n$, 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:

$$\begin{aligned} \vec{x} + \vec{y} &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \\ &= \begin{bmatrix} y_1 + x_1 \\ y_2 + x_2 \\ \vdots \\ y_n + x_n \end{bmatrix} \\ &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \vec{y} + \vec{x} \end{aligned} \tag{4.18.1}$$

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.

$$\begin{aligned} (\vec{x} + \vec{y}) + \vec{z} &= \left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right) + \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \\ &= \begin{bmatrix} (x_1 + y_1) + z_1 \\ (x_2 + y_2) + z_2 \\ \vdots \\ (x_n + y_n) + z_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 + (y_1 + z_1) \\ x_2 + (y_2 + z_2) \\ \vdots \\ x_n + (y_n + z_n) \end{bmatrix} \\ &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 + z_1 \\ y_2 + z_2 \\ \vdots \\ y_n + z_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \right) \\ &= \vec{x} + (\vec{y} + \vec{z}) \end{aligned} \tag{4.18.2}$$

Hence addition of vectors is associative.

- Next, we show the existence of an additive identity. Let $\vec{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

$$\begin{aligned} \vec{x} + \vec{0} &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} x_1 + 0 \\ x_2 + 0 \\ \vdots \\ x_n + 0 \end{bmatrix} \\ &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \vec{x} \end{aligned} \tag{4.18.3}$$

Hence the zero vector $\vec{0}$ is an additive identity.

- Next, we prove the existence of an additive inverse. Let $-\vec{x} = \begin{pmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{pmatrix}$.

$$\begin{aligned} \vec{x} + (-\vec{x}) &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 - x_1 \\ x_2 - x_2 \\ \vdots \\ x_n - x_n \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= \vec{0} \end{aligned} \tag{4.18.4}$$

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

$$
\begin{aligned}
a(\vec{x}+\vec{y}) &= a\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \\ \vdots x_n \end{bmatrix}\right) \\
&= a\begin{bmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{bmatrix} \\
&= \begin{bmatrix} a(x_1+y_1) \\ a(x_2+y_2) \\ \vdots \\ a(x_n+y_n) \end{bmatrix} \\
&= \begin{bmatrix} ax_1+ay_1 \\ ax_2+ay_2 \\ \vdots \\ ax_n+ay_n \end{bmatrix} \\
&= \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix} + \begin{bmatrix} ay_1 \\ ay_2 \\ \vdots \\ ay_n \end{bmatrix} \\
&= a\vec{x} + a\vec{y}
\end{aligned} \tag{4.18.5}
$$

- Next, we wish to show that $(a+b)\vec{x} = a\vec{x} + b\vec{x}$.

$$
\begin{aligned}
(a+b)\vec{x} &= (a+b)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} (a+b)x_1 \\ (a+b)x_2 \\ \vdots \\ (a+b)x_n \end{bmatrix} \\
&= \begin{bmatrix} ax_1+bx_1 \\ ax_2+bx_2 \\ \vdots \\ ax_n+bx_n \end{bmatrix} \\
&= \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix} + \begin{bmatrix} bx_1 \\ bx_2 \\ \vdots \\ bx_n \end{bmatrix} \\
&= a\vec{x} + b\vec{x}
\end{aligned} \tag{4.18.6}
$$

- We wish to show that $a(b\vec{x}) = (ab)\vec{x}$.

$$\begin{aligned} a(b\vec{x}) &= a\left(b\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) \\ &= a\left(\begin{bmatrix} bx_1 \\ bx_2 \\ \vdots \\ bx_n \end{bmatrix}\right) \\ &= \begin{bmatrix} a(bx_1) \\ a(bx_2) \\ \vdots \\ a(bx_n) \end{bmatrix} \\ &= \begin{bmatrix} (ab)x_1 \\ (ab)x_2 \\ \vdots \\ (ab)x_n \end{bmatrix} \\ &= (ab)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= (ab)\vec{x} \end{aligned} \tag{4.18.7}$$

- Finally, we need to show that $1\vec{x} = \vec{x}$.

$$\begin{aligned} 1\vec{x} &= 1\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \begin{bmatrix} 1x_1 \\ 1x_2 \\ \vdots \\ 1x_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \vec{x} \end{aligned} \tag{4.18.8}$$

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 4.18.2: 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

$$\mathbb{P}_2 = \left\{a_2x^2 + a_1x + a_0 | a_i^{\text{int}}\mathbb{R} \text{ for all } i\right\}$$

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.

$$\begin{aligned} p(x) + q(x) &= p_2x^2 + p_1x + p_0 + q_2x^2 + q_1x + q_0 \\ &= (p_2 + q_2)x^2 + (p_1 + q_1)x + (p_0 + q_0) \end{aligned} \tag{4.18.9}$$

  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 $p(x) + q(x) = q(x) + p(x)$.

$$\begin{aligned} p(x) + q(x) &= p_2x^2 + p_1x + p_0 + q_2x^2 + q_1x + q_0 \\ &= (p_2 + q_2)x^2 + (p_1 + q_1)x + (p_0 + q_0) \\ &= (q_2 + p_2)x^2 + (q_1 + p_1)x + (q_0 + p_0) \\ &= q_2x^2 + q_1x + q_0 + p_2x^2 + p_1x + p_0 \\ &= q(x) + p(x) \end{aligned} \tag{4.18.10}$$

- Next, we need to show that addition is associative. That is, that $(p(x) + q(x)) + r(x) = p(x) + (q(x) + r(x))$.

$$\begin{aligned} (p(x) + q(x)) + r(x) &= \left(p_2x^2 + p_1x + p_0 + q_2x^2 + q_1x + q_0\right) + r_2x^2 + r_1x + r_0 \\ &= (p_2 + q_2)x^2 + (p_1 + q_1)x + (p_0 + q_0) + r_2x^2 + r_1x + r_0 \\ &= (p_2 + q_2 + r_2)x^2 + (p_1 + q_1 + r_1)x + (p_0 + q_0 + r_0) \\ &= p_2x^2 + p_1x + p_0 + (q_2 + r_2)x^2 + (q_1 + r_1)x + (q_0 + r_0) \\ &= p_2x^2 + p_1x + p_0 + \left(q_2x^2 + q_1x + q_0 + r_2x^2 + r_1x + r_0\right) \\ &= p(x) + (q(x) + r(x)) \end{aligned} \tag{4.18.11}$$

- Next, we must prove that there exists an additive identity. Let $0(x) = 0x^2 + 0x + 0$.

$$\begin{aligned} p(x) + 0(x) &= p_2x^2 + p_1x + p_0 + 0x^2 + 0x + 0 \\ &= (p_2 + 0)x^2 + (p_1 + 0)x + (p_0 + 0) \\ &= p_2x^2 + p_1x + p_0 \\ &= p(x) \end{aligned} \tag{4.18.12}$$

  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:

$$\begin{aligned} p(x) + (-p(x)) &= p_2x^2 + p_1x + p_0 + \left(-p_2x^2 - p_1x - p_0\right) \\ &= (p_2 - p_2)x^2 + (p_1 - p_1)x + (p_0 - p_0) \\ &= 0x^2 + 0x + 0 \\ &= 0(x) \end{aligned} \tag{4.18.13}$$

  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 $a(p(x) + q(x)) = ap(x) + aq(x)$.

$$\begin{aligned} a(p(x) + q(x)) &= a\left(p_2x^2 + p_1x + p_0 + q_2x^2 + q_1x + q_0\right) \\ &= a\left((p_2 + q_2)x^2 + (p_1 + q_1)x + (p_0 + q_0)\right) \\ &= a(p_2 + q_2)x^2 + a(p_1 + q_1)x + a(p_0 + q_0) \\ &= (ap_2 + aq_2)x^2 + (ap_1 + aq_1)x + (ap_0 + aq_0) \\ &= ap_2x^2 + ap_1x + ap_0 + aq_2x^2 + aq_1x + aq_0 \\ &= ap(x) + aq(x) \end{aligned} \tag{4.18.14}$$

- Next we show that $(a + b)p(x) = ap(x) + bp(x)$.

$$\begin{aligned}(a+b)p(x) &= (a+b)(p_2x^2+p_1x+p_0)\\ &= (a+b)p_2x^2+(a+b)p_1x+(a+b)p_0\\ &= ap_2x^2+ap_1x+ap_0+bp_2x^2+bp_1x+bp_0\\ &= ap(x)+bp(x)\end{aligned} \tag{4.18.15}$$

- The next axiom which needs to be verified is $a(bp(x)) = (ab)p(x)$.

$$\begin{aligned}a(bp(x)) &= a\left(b\left(p_2x^2+p_1x+p_0\right)\right)\\ &= a\left(bp_2x^2+bp_1x+bp_0\right)\\ &= abp_2x^2+abp_1x+abp_0\\ &= (ab)(p_2x^2+p_1x+p_0)\\ &= (ab)p(x)\end{aligned} \tag{4.18.16}$$

- Finally, we show that $1p(x) = p(x)$.

$$\begin{aligned}1p(x) &= 1\left(p_2x^2+p_1x+p_0\right)\\ &= 1p_2x^2+1p_1x+1p_0\\ &= p_2x^2+p_1x+p_0\\ &= p(x)\end{aligned} \tag{4.18.17}$$

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 4.18.3: 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.

$$\begin{aligned}A+B &= \begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\end{bmatrix}+\begin{bmatrix}b_{11} & b_{12} & b_{13}\\ b_{21} & b_{22} & b_{23}\end{bmatrix}\\ &= \begin{bmatrix}a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13}\\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23}\end{bmatrix}\end{aligned} \tag{4.18.18}$$

  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.

$$\begin{aligned}cA &= c\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\end{bmatrix}\\ &= \begin{bmatrix}ca_{11} & ca_{12} & ca_{13}\\ ca_{21} & ca_{22} & ca_{23}\end{bmatrix}\end{aligned} \tag{4.18.19}$$

  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 4.18.4: 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

$$\begin{aligned} A + B &= A \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ B + A &= B \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{aligned} \tag{4.18.20}$$

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 4.18.5: 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

$$\begin{aligned} (f + g)(x) &= f(x) + g(x) \\ (af)(x) &= a(f(x)) \end{aligned} \tag{4.18.21}$$

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.

One interseting example is to find the zero elment and the addivtie invse for the first set in the tabvle

Example 4.18.1

Finding the Zero Element and Additive Inverse

Let the zero element be $\vec{0}=(a,b)$. To satisfy the zero property, we must have:

$$(x,y)+(a,b)=(x,y) \tag{4.18.22}$$

Using the given operation:

$$(x,y)+(a,b)=(x+a+1,\ y+b) \tag{4.18.23}$$

Set the result equal to (x,y):

$$(x+a+1,\ y+b)=(x,y) \tag{4.18.24}$$

Comparing components gives:

$$a+1=0 \quad \text{and} \quad b=0 \tag{4.18.25}$$

Therefore, the zero element is:

$$\vec{0}=(-1,\ 0) \tag{4.18.26}$$

--- Now, let the additive inverse of (x,y) be (u,v) such that:

$$(x, y) + (u, v) = \vec{0} \tag{4.18.27}$$

Substitute the operation:

$$(x + u + 1,\ y + v) = (-1,\ 0) \tag{4.18.28}$$

From this we get:

$$u = -x - 2 \quad \text{and} \quad v = -y \tag{4.18.29}$$

Hence, the additive inverse of (x, y) is:

$$-(x, y) = (-x - 2,\ -y) \tag{4.18.30}$$

Consider the following important theorem.

Theorem 4.18.1: 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 available at the end of this section.

An important use of the additive inverse is the following theorem.

Theorem 4.18.2

Let V be a vector space. Then $\vec{v} + \vec{w} = \vec{v} + \vec{z}$ implies that $\vec{w} = \vec{z}$ for all $\vec{v}, \vec{w}, \vec{z}\,^{\text{int}} V$

Proof available at the end of this section.

Exercise

Exercise 4.18.1

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 4.18.2

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 4.18.3

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 4.18.4

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 4.18.5

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

$$\begin{aligned} (f+g)(x) &= f(x) + g(x) \\ (af)(x) &= a(f(x)) \end{aligned} \tag{4.18.31}$$

Exercise 4.18.6

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 4.18.5? Is this a vector space?

Exercise 4.18.7

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 4.18.8

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 4.18.9

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 4.18.10

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.

Answers

Theorem 4.18.1**: Uniqueness**

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

   $$\begin{aligned} 0\vec{x} + (-0\vec{x}) &= 0\vec{x} + 0\vec{x} + (-\vec{x}) \\ \vec{0} + 0 &= 0\vec{x} + 0 \\ \vec{0} &= 0\vec{x} \end{aligned} \tag{4.18.32}$$

   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:

   $$\begin{aligned} (-1)\,\vec{x} + \vec{x} &= (-1)\,\vec{x} + 1\vec{x} \\ &= (-1 + 1)\,\vec{x} \\ &= 0\vec{x} \\ &= \vec{0} \end{aligned} \tag{4.18.33}$$

   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 $(-1)\,\vec{x} = -\vec{x}$.

Theorem 4.18.2

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 4.18: Amira Vector Space Axioms is shared under a license and was authored, remixed, and/or curated by LibreTexts.

- 9.1: Algebraic Considerations by Ken Kuttler is licensed CC BY 4.0. Original source: https://lyryx.com/first-course-linear-algebra.
- 21.2: Vector Spaces by Dirk Colbry is licensed CC BY-NC 4.0. Original source: https://colbrydi.github.io/MatrixAlgebra.

4.19: Amira Vector Subspace of Rⁿ

Learning Objectives

1. Definition of a subspace.
2. Determine whether or not a subset is a subspace.
3. Learn the most important examples of subspaces.
4. Write a given subspace as a column space or null space.
5. Compute a spanning set for a null space.
6. Visualize subset of $\mathbb{R}^2$ or $\mathbb{R}^3$ is a subspace or not.

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.

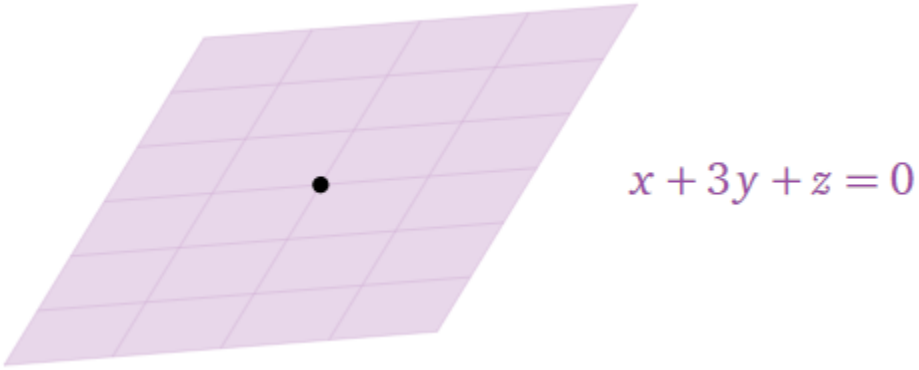


Figure 4.19.1

(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 4.19.1: Subset

A ***subset*** of $\mathbb{R}^n$ is any collection of points of $\mathbb{R}^n$.
Let U be a set of vectors in $\mathbb{R}^n$. Then we say that U is a **subset** of $\mathbb{R}^n$, denoted

$$U \subset \mathbb{R}^n$$

For instance, the unit circle

$$C = \left\{(x, y) \text{ in } \mathbb{R}^2 \mid x^2 + y^2 = 1\right\}$$

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 4.19.2: 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 $\mathrm{Span}\{u, v\}$ is contained in V
- Similarly, if $v_1, v_2, \ldots, v_n$ are all in V, then $\mathrm{Span}\{v_1, v_2, \ldots, v_n\}$ 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 4.19.1 below for a precise statement.

Remark

A quick test to determine whether a subset $V \subseteq \mathbb{R}^n$ is a subspace is:

1. The zero vector 0 is in V.
2. For any vectors v_1, v_2 in V and any scalars c, d, the linear combination

$$cv_1 - dv_2 \text{ in } V. \tag{4.19.1}$$

Example 4.19.1

The set $\mathbb{R}^n$ is a subspace of itself: indeed, it contains zero, and is closed under addition and scalar multiplication.

Example 4.19.2

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 4.19.3: A line through the origin

A line L through the origin is a subspace.

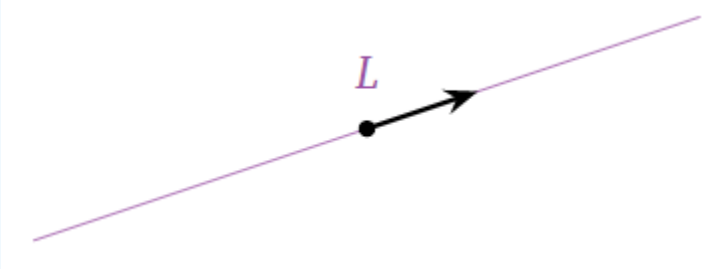


Figure 4.19.2

Indeed, L contains zero, and is easily seen to be closed under addition and scalar multiplication.

Example 4.19.4: A plane through the origin

A plane P through the origin is a subspace.

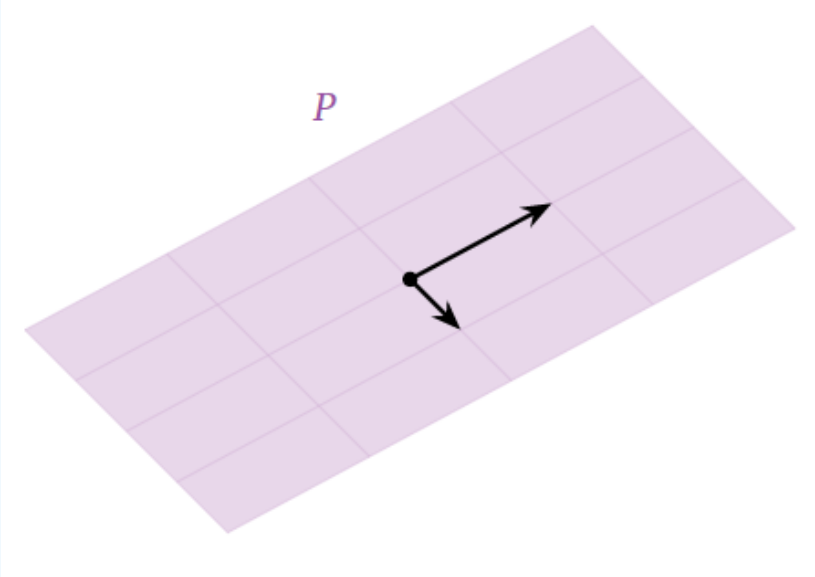


Figure 4.19.3

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 4.19.5: 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.

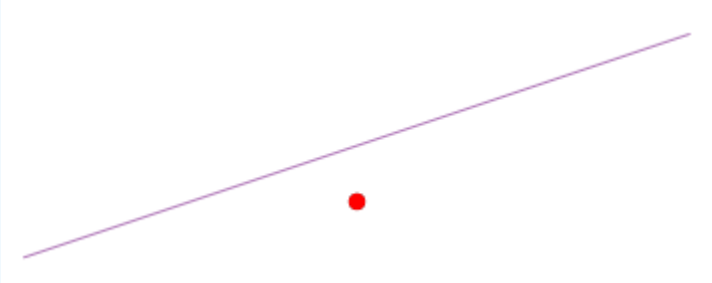


Figure 4.19.4

Example 4.19.6: 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.

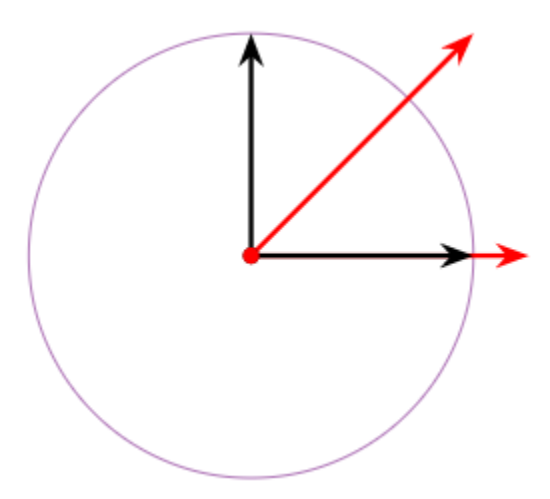


Figure 4.19.5

Example 4.19.7: 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).

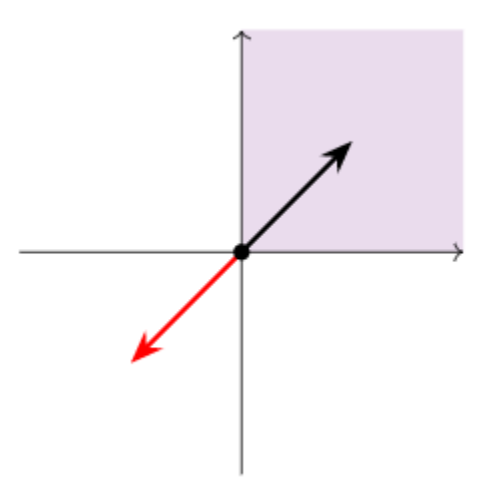


Figure 4.19.6

Example 4.19.8: 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.

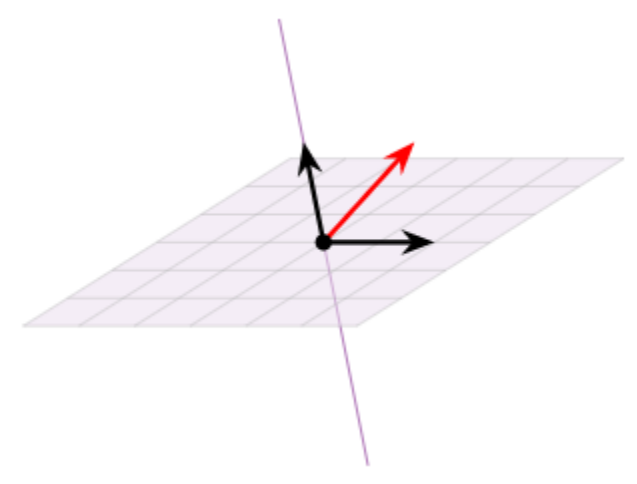


Figure 4.19.7

Note 4.19.1: Subsets versus Subspaces

A subs*et* of $\mathbb{R}^n$ is any collection of vectors whatsoever. For instance, the unit circle

$$C = \left\{(x, y) \text{ in } \mathbb{R}^2 \mid x^2 + y^2 = 1\right\}$$

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 4.19.3, below.

Example 4.19.9: 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 $\begin{pmatrix} a \\ b \end{pmatrix}$ 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, $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 1/2 \\ 1/3 \end{pmatrix}$ are in V, but $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ is not because $2 \cdot 2 \neq 3 \cdot 3$.

Let us check the first property. The subset V does contain the zero vector $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, 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 $u = \begin{pmatrix} a \\ b \end{pmatrix}$ and $v = \begin{pmatrix} c \\ d \end{pmatrix}$ 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 $\begin{pmatrix} a+c \\ b+d \end{pmatrix}$ 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 $v = \begin{pmatrix} a \\ b \end{pmatrix}$ 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 $\begin{pmatrix} ca \\ cb \end{pmatrix}$ 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 $2/3$.

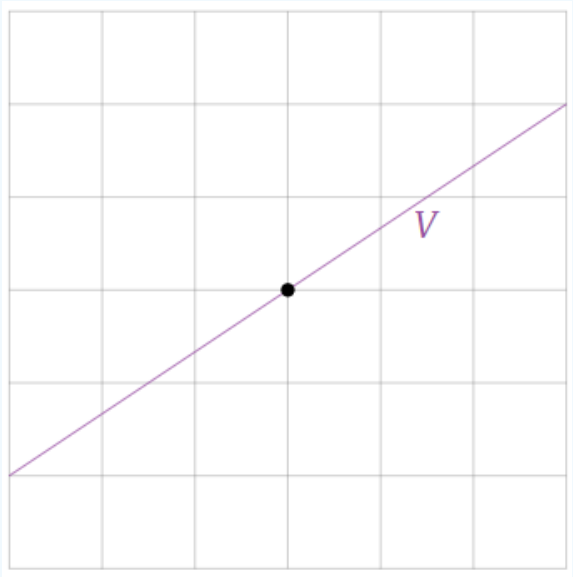


Figure 4.19.8

Example 4.19.10: 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.

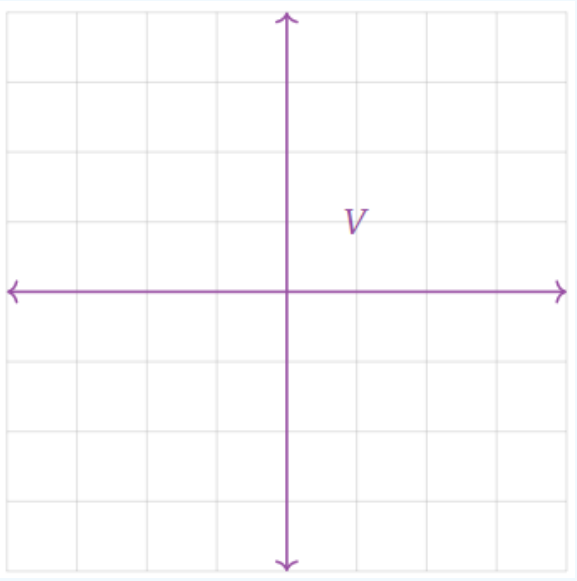


Figure 4.19.9

Common Types of Subspaces

Definition: Linear Combination of vectors

Let $v_1, v_2, \dots, v_k$ be vectors in a vector space V, and let
$c_1, c_2, \dots, c_k$ be scalars.
A vector of the form

$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k \tag{4.19.2}$$

is called a **linear combination** of the vectors $v_1, v_2, \dots, v_k$.

If we are given a set of vectors, we can create infinitely many vectors from the linear combination formal.

Example 4.19.1

Consider the vectors

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix}, \qquad \mathbf{v}_3 = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}. \tag{4.19.3}$$

A linear combination of these vectors is obtained by choosing scalars
$c_1 = 2$, $c_2 = -1$, and $c_3 = 4$, giving

$$2\mathbf{v}_1 - \mathbf{v}_2 + 4\mathbf{v}_3 = 2\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} - \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix} + 4\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 11 \\ 0 \\ 1 \end{pmatrix}. \tag{4.19.4}$$

Thus,

$$\mathbf{w}\begin{pmatrix} 11 \\ 0 \\ 1 \end{pmatrix} \tag{4.19.5}$$

is a linear combination of $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$.

The set that includes all of these vectors is defining bellow.

Definition 4.19.2: 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 4.19.1: Span of Vectors

Describe the span of the vectors $\vec{u} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T$ and Double exponent: use braces to clarify .

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. This concept is explained below in a interactive video format. Feel free to click on it and see the theory in action.

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 $\vec{v}$.

Theorem 4.19.1: 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 available at the end of this section.

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 4.19.3: 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 4.19.1. 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 $\mathrm{Nul}(A)$ because $A0 = 0$.
2. Suppose that u, v are in $\mathrm{Nul}(A)$. 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 $\mathrm{Nul}(A)$.
3. Suppose that v is in $\mathrm{Nul}(A)$ 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 $\mathrm{Nul}(A)$.

Since $\mathrm{Nul}(A)$ satisfies the three defining properties of a subspace, it is a subspace.

Example 4.19.11

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:
$$\mathrm{Col}(A) = \mathrm{Span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\} = \mathrm{Span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

This is a line in $\mathbb{R}^3$.

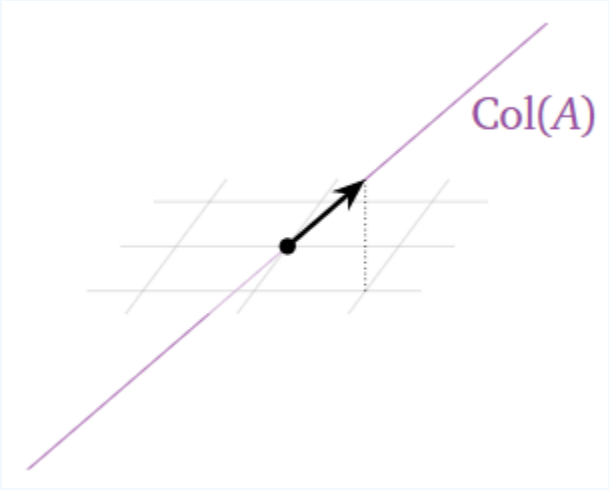


Figure 4.19.10

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 $\mathbb{R}^2$.

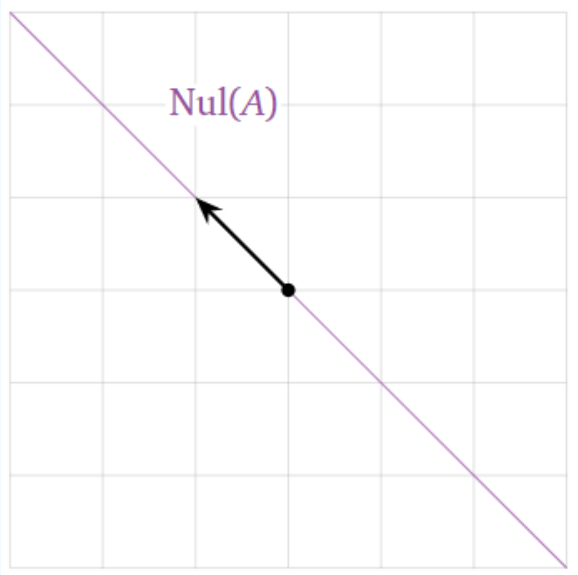


Figure 4.19.11

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 4.19.2: 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 4.19.12: 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 4.19.13: 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 4.19.3

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 4.19.14

Let

$$V = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \text{ in } \mathbb{R}^2 | 2a = 3b \right\}$$

be the subset of a previous Example 4.19.9. The subset V is exactly the solution set of the homogeneous equation $2x - 3y = 0$. Therefore,

$$V = \text{Nul}(2 \; -3)$$

In particular, it is a subspace. The reduced row echelon form of $\begin{pmatrix} 2 & -3 \end{pmatrix}$ is $\begin{pmatrix} 1 & -3/2 \end{pmatrix}$, 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 4.19.15

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

$$V = \text{Col}\begin{pmatrix} 2 & 1 \\ 1 & -1 \\ 3 & -2 \end{pmatrix}.$$

Exercises

Exercise 4.19.6

Let $H = \text{span}\left\{\begin{bmatrix} 0 \\ 2 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 6 \\ 0 \\ -2 \end{bmatrix}, \begin{bmatrix} -2 \\ 16 \\ 0 \\ -6 \end{bmatrix}, \begin{bmatrix} -3 \\ 22 \\ 0 \\ -8 \end{bmatrix}\right\}$. Find the dimension of H and determine a basis.

Exercise 4.19.7

Let $H = \text{span}\left\{\begin{bmatrix} 5 \\ 1 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 14 \\ 3 \\ 2 \\ 8 \end{bmatrix}, \begin{bmatrix} 38 \\ 8 \\ 6 \\ 24 \end{bmatrix}, \begin{bmatrix} 47 \\ 10 \\ 7 \\ 28 \end{bmatrix}, \begin{bmatrix} 10 \\ 2 \\ 3 \\ 12 \end{bmatrix}\right\}$. Find the dimension of H and determine a basis.

Exercise 4.19.8

Let $H = \text{span}\left\{\begin{bmatrix} 6 \\ 1 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 17 \\ 3 \\ 2 \\ 10 \end{bmatrix}, \begin{bmatrix} 52 \\ 9 \\ 7 \\ 35 \end{bmatrix}, \begin{bmatrix} 18 \\ 3 \\ 4 \\ 20 \end{bmatrix}\right\}$. Find the dimension of H and determine a basis.

Exercise 4.19.9

Let $M = \left\{\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}^{\text{int}} \mathbb{R}^4 : \sin(u_1) = 1\right\}$. Is M a subspace? Explain.

Answer available at the end of this section.

Exercise 4.19.10

Let $M = \left\{\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}^{\text{int}} \mathbb{R}^4 : ||u_1|| \leq 4\right\}$. Is M a subspace? Explain.

Answer available at the end of this section.

Exercise 4.19.11

Let $M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}^{\text{int}} \mathbb{R}^4 : u_1 \geq 0 \text{ for each } i = 1, 2, 3, 4 \right\}$. Is M a subspace? Explain.

Answer available at the end of this section.

Exercise 4.19.12

Let $\vec{w}$, $\vec{w}_1$ be given vectors in $\mathbb{R}^4$ and define

$$M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}^{\text{int}} \mathbb{R}^4 : \vec{w} \bullet \vec{u} = 0 \text{ and } \vec{w}_1 \bullet \vec{u} = 0 \right\}.$$

Is M a subspace? Explain.

Answer available at the end of this section.

Exercise 4.19.13

Let $\vec{w}^{\text{int}}\mathbb{R}^4$ and let $M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}^{\text{int}} \mathbb{R}^4 : \vec{w} \bullet \vec{u} = 0 \right\}$. Is M a subspace? Explain.

Answer available at the end of this section.

Exercise 4.19.14

Let $M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}^{\text{int}} \mathbb{R}^4 : u_3 \geq u_1 \right\}$. Is M a subspace? Explain.

Answer available at the end of this section.

Exercise 4.19.15

Let $M = \left\{ \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}^{\text{int}} \mathbb{R}^4 : u_3 = u_1 = 0 \right\}$. Is M a subspace? Explain.

Answer available at the end of this section.

Exercise 4.19.16

Consider the set of vectors S given by

$$S = \left\{ \left[\begin{array}{c} 4u+v-5w \\ 12u+6v-6w \\ 4u+4v+4w \end{array} \right] : u,v,w \in \mathbb{R} \right\}.$$

Is S a subspace of $\mathbb{R}^3$?

Exercise 4.19.17

Consider the set of vectors S given by

$$S = \left\{ \left[\begin{array}{c} 2u+6v+7w \\ -3u-9v-12w \\ 2u+6v+6w \\ u+3v+3w \end{array} \right] : u,v,w \in \mathbb{R} \right\}.$$

Is S a subspace of $\mathbb{R}^4$?

Exercise 4.19.18

Consider the set of vectors S given by

$$S = \left\{ \left[\begin{array}{c} 2u+v \\ 6v-3u+3w \\ 3v-6u+3w \end{array} \right] : u,v,w \in \mathbb{R} \right\}.$$

Is this set of vectors a subspace of $\mathbb{R}^3$?

Exercise 4.19.19

Consider the vectors of the form

$$\left\{ \left[\begin{array}{c} 2u+v+7w \\ u-2v+w \\ -6v-6w \end{array} \right] : u,v,w \in \mathbb{R} \right\}.$$

Is this set of vectors a subspace of $\mathbb{R}^3$?

Exercise 4.19.20

Consider the vectors of the form

$$\left\{ \left[\begin{array}{c} 3u+v+11w \\ 18u+6v+66w \\ 28u+8v+100w \end{array} \right] : u,v,w \in \mathbb{R} \right\}.$$

Is this set of vectors a subspace of $\mathbb{R}^3$?

Exercise 4.19.21

Consider the vectors of the form

$$\left\{ \begin{bmatrix} 3u+v \\ 2w-4u \\ 2w-2v-8u \end{bmatrix} : u,v,w^{\text{int}}\mathbb{R} \right\}.$$

Is this set of vectors a subspace of $\mathbb{R}^3$?

Exercise 4.19.22

Consider the set of vectors S given by

$$\left\{ \begin{bmatrix} u+v+w \\ 2u+2v+4w \\ u+v+w \\ 0 \end{bmatrix} : u,v,w^{\text{int}}\mathbb{R} \right\}.$$

Is S is a subspace of $\mathbb{R}^4$?

Exercise 4.19.23

Consider the set of vectors S given by

$$\left\{ \begin{bmatrix} v \\ -3u-3w \\ 8u-4v+4w \end{bmatrix} : u,v,w^{\text{int}}\mathbb{R} \right\}.$$

Is S is a subspace of $\mathbb{R}^4$?

Exercise 4.19.24

If you have 5 vectors in $\mathbb{R}^5$ and the vectors are linearly independent, can it always be concluded they span $\mathbb{R}^5$? Explain.

Answer available at the end of this section.

Exercise 4.19.25

If you have 6 vectors in $\mathbb{R}^5$, is it possible they are linearly independent? Explain.

Answer available at the end of this section.

Exercise 4.19.26

Suppose A is an $m \times n$ matrix and $\{\vec{w}_1, \cdots, \vec{w}_k\}$ is a linearly independent set of vectors in $A(\mathbb{R}^n) \subseteq \mathbb{R}^m$. Now suppose $A\vec{z}_i = \vec{w}_i$. Show $\{\vec{z}_1, \cdots, \vec{z}_k\}$ is also independent.

Answer available at the end of this section.

Exercise 4.19.27

Let V and W be subspaces of $\mathbb{R}^n$. Show that the intersection $V \cap W$ is also a subspace of $\mathbb{R}^n$

Answers

Theorem 4.19.1**: Spans are Subspaces and Subspaces are Spans**

To show that $\text{Span}\{v_1, v_2, \dots, v_p\}$ 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 $\text{Span}\{v_1, v_2, \dots, v_p\}$, 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 $\text{Span}\{v_1, v_2, \dots, v_p\}$.
3. If $v = a_1v_1 + a_2v_2 + \cdots + a_pv_p$ is in $\text{Span}\{v_1, v_2, \dots, v_p\}$ and c is a scalar, then
$$cv = ca_1v_1 + ca_2v_2 + \cdots + ca_pv_p$$
is also in $\text{Span}\{v_1, v_2, \dots, v_p\}$.

Since $\text{Span}\{v_1, v_2, \dots, v_p\}$ 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 $V = \text{Span}\{v_1, v_2, \dots, v_p\}$ for some linearly independent set $\{v_1, v_2, \dots, v_p\}$. This process terminates after at most n steps by this important note in Section 2.5, Note 2.5.2.

Exercise 4.19.9

No. Let $\vec{u} = \begin{bmatrix} \frac{\pi}{2} \\ 0 \\ 0 \\ 0 \end{bmatrix}$. Then $2\vec{u} \overset{\text{int}}{\not\in} M$ although $\vec{u} \overset{\text{int}}{\in} M$.

Exercise 4.19.10

No. $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \overset{\text{int}}{\in} M$ but $10 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \overset{\text{int}}{\not\in} M$.

Exercise 4.19.11

This is not a subspace. $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ is in it. However, $(-1) \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ is not.

Exercise 4.19.12

This is a subspace because it is closed with respect to vector addition and scalar multiplication.

Exercise 4.19.13

Yes, this is a subspace because it is closed with respect to vector addition and scalar multiplication.

Exercise 4.19.14

This is not a subspace. $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ is in it. However $(-1) \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}$ is not.

Exercise 4.19.15

This is a subspace. It is closed with respect to vector addition and scalar multiplication.

Exercise 4.19.24

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 4.19.25

They can't be.

Exercise 4.19.26

Say $\sum_{i=1}^{k} c_i \vec{z}_i = \vec{0}$. Then apply A to it as follows.

$$\sum_{i=1}^{k} c_a A\vec{z}_i = \sum_{i=1}^{k} c_i \vec{w}_i = \vec{0}$$

and so, by linear independence of the $\vec{w}_i$, it follows that each $c_i = 0$.

This page titled 4.19: Amira Vector Subspace of R^n is shared under a license and was authored, remixed, and/or curated by LibreTexts.

- 2.6: Subspaces by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.
- 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.

4.20: The Real Vector Space ^R^n

This page titled 4.20: The Real Vector Space ^R^n is shared under a license and was authored, remixed, and/or curated by LibreTexts.

4.E: Exercises

Exercise 4.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 4.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 4.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 4.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 4.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

$$\begin{aligned}(f+g)(x)&=f(x)+g(x)\\(af)(x)&=a(f(x))\end{aligned}\tag{4.E.1}$$

Exercise 4.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 4.E.5? Is this a vector space?

Exercise $4.E.7$

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 $4.E.8$

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 $4.E.9$

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 $4.E.10$

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 $4.E.11$

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 $4.E.12$

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 $4.E.13$

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 $4.E.14$

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 $4.E.15$

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 $4.E.16$

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 $4.E.17$

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 $4.E.18$

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 $4.E.19$

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 $4.E.20$

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 available at the end of this section.

Exercise $4.E.21$

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 available at the end of this section.

Exercise $4.E.22$

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 available at the end of this section.

Exercise $4.E.23$

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 available at the end of this section.

Exercise $4.E.24$

Determine if $p(x) = 4x^2 - x$ is in the span given by

$$span\{x^2 + x,\ x^2 - 1,\ -x + 2\}$$

Exercise $4.E.25$

Determine if $p(x) = -x^2 + x + 2$ is in the span given by

$$span\{x^2 + x + 1,\ 2x^2 + x\}$$

Exercise $4.E.26$

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 $4.E.27$

Show that the spanning set in Exercise $4.E.26$ is a spanning set for M_{22}, the vector space of all 2×2 matrices.

Exercise $4.E.28$

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 $4.E.29$

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 $4.E.30$

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 $4.E.31$

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 $4.E.32$

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 $4.E.33$

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 $4.E.34$

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 $4.E.35$

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 $4.E.36$

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 $4.E.37$

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 $4.E.38$

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 $4.E.39$

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 $4.E.40$

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 $4.E.41$

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 $4.E.42$

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 $4.E.43$

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 $4.E.44$

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 $4.E.45$

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 $4.E.46$

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 $4.E.47$

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 $4.E.48$

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 $4.E.49$

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 $4.E.50$

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 $4.E.51$

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 $4.E.52$

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 $4.E.53$

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 $4.E.54$

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 $4.E.55$

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 $4.E.56$

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 available at the end of this section.

Exercise $4.E.57$

If you have 6 vectors in $\mathbb{R}^5$, is it possible they are linearly independent? Explain.

Answer available at the end of this section.

Exercise $4.E.58$

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 available at the end of this section.

Exercise $4.E.59$

In the context of the above problem, consider polynomials

$$\{a_i x^3+b_i x^2+c_i x+d_i,\ i=1,2,3,4\}$$

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 available at the end of this section.

Exercise $4.E.60$

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 available at the end of this section.

Exercise $4.E.61$

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 available at the end of this section.

Exercise $4.E.62$

Let $M = \{\vec{u} = (u_1,\ u_2,\ u_3,\ u_4)^{\text{int}}\mathbb{R}^4\ :\ |u_1| \leq 4\}$. Is M a subspace of $\mathbb{R}^4$?

Answer available at the end of this section.

Exercise $4.E.63$

Let $M = \{\vec{u} = (u_1,\ u_2,\ u_3,\ u_4)^{\text{int}}\mathbb{R}^4\ :\ \sin(u_1) = 1\}$. Is M a subspace of $\mathbb{R}^4$?

Answer available at the end of this section.

Exercise $4.E.64$

Let W be a subset of M_{22} given by

$$W = \{A | A^{\text{int}}M_{22}, A^T = A\}$$

In words, W is the set of all symmetric 2×2 matrices. Is W a subspace of M_{22}?

Exercise $4.E.65$

Let W be a subset of M_{22} given by

$$W = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} | a, b, c, d^{\text{int}}\mathbb{R},\ a + b = c + d \right\}$$

Is W a subspace of M_{22}?

Exercise $4.E.66$

Let W be a subset of P_3 given by

$$W = \{ax^3 + bx^2 + cx + d |\ a, b, c, d^{\text{int}}\mathbb{R}, d = 0\}$$

Is W a subspace of P_3?

Exercise $4.E.67$

Let W be a subset of P_3 given by

$$W = \{p(x) = ax^3 + bx^2 + cx + d |\ a, b, c, d^{\text{int}}\mathbb{R}, p(2) = 1\}$$

Is W a subspace of P_3?

Exercise $4.E.68$

Let $T\colon \mathbb{P}_2 \to \mathbb{R}$ 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 available at the end of this section.

Exercise $4.E.69$

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. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x + 2y^2 + 3z \\ 2y + 3z + z \end{bmatrix}$

c. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \sin x + 2y + 3z \\ 2y + 3z + z \end{bmatrix}$

d. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x + 2y + 3z \\ 2y + 3z - \ln z \end{bmatrix}$

Answer available at the end of this section.

Exercise $4.E.70$

Suppose T is a linear transformation such that

$$\begin{aligned} T\begin{bmatrix} 1 \\ 1 \\ -7 \end{bmatrix} &= \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} \\ T\begin{bmatrix} -1 \\ 0 \\ 6 \end{bmatrix} &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \\ T\begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} &= \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} \end{aligned} \tag{4.E.2}$$

Find the matrix of T. That is find A such that $T(\vec{x}) = A\vec{x}$.

Answer available at the end of this section.

Exercise $4.E.71$

Suppose T is a linear transformation such that

$$T\begin{bmatrix} 1 \\ 2 \\ -18 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 5 \end{bmatrix}$$
$$T\begin{bmatrix} -1 \\ -1 \\ 15 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 5 \end{bmatrix} \tag{4.E.3}$$
$$T\begin{bmatrix} 0 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ -2 \end{bmatrix}$$

Find the matrix of T. That is find A such that $T(\vec{x}) = A\vec{x}$.

Answer available at the end of this section.

Exercise $4.E.72$

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. $T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2y - 5x + z \\ x + y + z \end{bmatrix}$

Answer available at the end of this section.

Exercise $4.E.73$

Suppose

$$[A_1 \cdots A_n]^{-1}$$

exists where each $A_j^{\text{int}}\mathbb{R}^n$ 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 $4.E.74$

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 $\{\vec{v}_1, \cdots, \vec{v}_r\}$ is also linearly independent.

Answer available at the end of this section.

Exercise $4.E.75$

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 $4.E.76$

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 available at the end of this section.

Exercise $4.E.77$

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 $4.E.78$

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 $span\{T\vec{v}_1, \cdots, T\vec{v}_r\} = W$.

Exercise $4.E.79$

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 $4.E.80$

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 $4.E.81$

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 $4.E.82$

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

Exercise $4.E.83$

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 $4.E.84$

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 $4.E.85$

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 $4.E.86$

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 available at the end of this section.

Exercise $4.E.87$

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 available at the end of this section.

Exercise $4.E.88$

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 available at the end of this section.

Exercise $4.E.89$

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 available at the end of this section.

Exercise $4.E.90$

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 available at the end of this section.

Exercise $4.E.91$

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 ih 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 available at the end of this section.

Exercise $4.E.92$

You are given a linear transformation T: $\mathbb{R}^n \to \mathbb{R}^m$ and you know that

$$T(A_i) = B_i$$

where $\begin{bmatrix} A_1 & \cdots & A_n \end{bmatrix}^{-1}$ exists. Show that the matrix of T is of the form

$$\begin{bmatrix} B_1 & \cdots & B_n \end{bmatrix} \begin{bmatrix} A_1 & \cdots & A_n \end{bmatrix}^{-1}$$

Answer available at the end of this section.

Exercise $4.E.93$

Suppose T is a linear transformation such that

$$\begin{aligned} T\begin{bmatrix} 1 \\ 2 \\ -6 \end{bmatrix} &= \begin{bmatrix} 5 \\ 1 \\ 3 \end{bmatrix} \\ T\begin{bmatrix} -1 \\ -1 \\ 5 \end{bmatrix} &= \begin{bmatrix} 1 \\ 1 \\ 5 \end{bmatrix} \\ T\begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} &= \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix} \end{aligned} \tag{4.E.4}$$

Find the matrix of T. That is find A such that $T(\vec{x}) = A\vec{x}$.

Answer available at the end of this section.

Exercise $4.E.94$

Suppose T is a linear transformation such that

$$T\begin{bmatrix} 1 \\ 1 \\ -8 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$$
$$T\begin{bmatrix} -1 \\ 0 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} \quad (4.E.5)$$
$$T\begin{bmatrix} 0 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \\ -1 \end{bmatrix}$$

Find the matrix of T. That is find A such that $T(\vec{x}) = A\vec{x}$.

Answer available at the end of this section.

Exercise $4.E.95$

Suppose T is a linear transformation such that

Illegal pream-token (4)

Find the matrix of T. That is find A such that $T(\vec{x}) = A\vec{x}$.

Answer available at the end of this section.

Exercise $4.E.96$

Suppose T is a linear transformation such that

$$T\begin{bmatrix} 1 \\ 1 \\ -7 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$$
$$T\begin{bmatrix} -1 \\ 0 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad (4.E.6)$$
$$T\begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}$$

Find the matrix of T. That is find A such that $T(\vec{x}) = A\vec{x}$.

Answer available at the end of this section.

Exercise $4.E.97$

Suppose T is a linear transformation such that

$$T\begin{bmatrix}1\\2\\-18\end{bmatrix}=\begin{bmatrix}5\\2\\5\end{bmatrix}$$
$$T\begin{bmatrix}-1\\-1\\15\end{bmatrix}=\begin{bmatrix}3\\3\\5\end{bmatrix} \quad (4.E.7)$$
$$T\begin{bmatrix}0\\-1\\4\end{bmatrix}=\begin{bmatrix}2\\5\\-2\end{bmatrix}$$

Find the matrix of T. That is find A such that $T(\vec{x})=A\vec{x}$.

Answer available at the end of this section.

Exercise $4.E.98$

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. $T\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}2y-5x+z\\x+y+z\end{bmatrix}$

Answer available at the end of this section.

Exercise $4.E.99$

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. $T\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}x+2y+3z\\2y+3x-\ln z\end{bmatrix}$

Answer available at the end of this section.

Exercise 4.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 4.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 available at the end of this section.

Exercise 4.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 available at the end of this section.

Exercise 4.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 available at the end of this section.

Exercise 4.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 4.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 available at the end of this section.

Exercise 4.E.106

Let $T\colon \mathbb{R}^2 \mapsto \mathbb{R}^2$ 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 available at the end of this section.

Answers

Exercise $4.E.20$

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 $4.E.21$

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 $4.E.22$

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 $4.E.23$

$\sum_{i=1}^{k} 0\vec{x}_k = \vec{0}$

Exercise $4.E.56$

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 $4.E.57$

No. They can't be.

Exercise $4.E.58$

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

$$\begin{aligned} c_1 + 2c_3 + 2c_4 &= 0 \\ c_2 + c_3 - c_4 &= 0 \\ c_2 - 3c_4 &= 0 \\ c_1 + c_4 &= 0 \end{aligned} \tag{4.E.8}$$

, Solution is:

$$[c_1 = 0,\ c_2 = 0,\ c_3 = 0,\ c_4 = 0]$$

Therefore, these are linearly independent.

Exercise $4.E.59$

Let $p_i(x)$ denote the ith of these polynomials. Suppose $\sum_i C_i p_i(x) = 0$. Then collecting terms according to the exponent of x, you need to have

$$\begin{aligned} C_1a_1 + C_2a_2 + C_3a_3 + C_4a_4 &= 0 \\ C_1b_1 + C_2b_2 + C_3b_3 + C_4b_4 &= 0 \\ C_1c_1 + C_2c_2 + C_3c_3 + C_4c_4 &= 0 \\ C_1d_1 + C_2d_2 + C_3d_3 + C_4d_4 &= 0 \end{aligned} \tag{4.E.9}$$

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 $4.E.60$

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 $4.E.61$

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 $4.E.62$

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 $4.E.63$

This is not a subspace.

Exercise $4.E.68$

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 $T(ax^2+bx+c)=aT(x^2)+bT(x)+cT(1)=a+5b-6c$.

Exercise $4.E.69$

Exercise $4.E.70$

$$\begin{bmatrix}3&1&1\\3&2&3\\3&3&-1\end{bmatrix}\begin{bmatrix}6&2&1\\5&2&1\\6&1&1\end{bmatrix}=\begin{bmatrix}29&9&5\\46&13&8\\27&11&5\end{bmatrix}$$

Exercise $4.E.71$

$$\begin{bmatrix}5&3&2\\2&3&5\\5&5&-2\end{bmatrix}\begin{bmatrix}11&4&1\\10&4&1\\12&3&1\end{bmatrix}=\begin{bmatrix}109&38&10\\112&35&10\\81&34&8\end{bmatrix}$$

Exercise $4.E.72$

Exercise $4.E.74$

If $\sum_i^r a_i\vec{v}_r=0$, then using linearity properties of T we get

$$0=T(0)=T\left(\sum_i^r a_i\vec{v}_r\right)=\sum_i^r a_iT(\vec{v}_r).$$

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 $4.E.76$

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 $4.E.86$

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 $4.E.87$

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 $4.E.88$

In this case $\ker(T)=\{0\}$ and $Im(T)=\mathbb{R}^2$ (pick any basis of $\mathbb{R}^2$).

Exercise $4.E.89$

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 $4.E.90$

We can easily see that $\dim(Im(T))=1$, and thus $\dim(\ker(T))=3-\dim(Im(T))=3-1=2$.

Exercise $4.E.91$

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 $4.E.92$

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

$$\begin{aligned}\begin{bmatrix}\vec{b}_1 & \cdots & \vec{b}_n\end{bmatrix}\begin{bmatrix}\vec{a}_1 & \cdots & \vec{a}_n\end{bmatrix}^{-1}\vec{a}_i &= \begin{bmatrix}\vec{b}_1 & \cdots & \vec{b}_n\end{bmatrix}\begin{bmatrix}\vec{c}_1^T \\ \vdots \\ \vec{c}_n^T\end{bmatrix}\vec{a}_i \\ &= \begin{bmatrix}\vec{b}_1 & \cdots & \vec{b}_n\end{bmatrix}\vec{e}_i \\ &= \vec{b}_i\end{aligned} \tag{4.E.10}$$

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 $4.E.93$

$$\begin{bmatrix}5 & 1 & 5\\ 1 & 1 & 3\\ 3 & 5 & -2\end{bmatrix}\begin{bmatrix}3 & 2 & 1\\ 2 & 2 & 1\\ 4 & 1 & 1\end{bmatrix} = \begin{bmatrix}37 & 17 & 11\\ 17 & 7 & 5\\ 11 & 14 & 6\end{bmatrix}$$

Exercise $4.E.94$

$$\begin{bmatrix}1 & 2 & 6\\ 3 & 4 & 1\\ 1 & 1 & -1\end{bmatrix}\begin{bmatrix}6 & 3 & 1\\ 5 & 3 & 1\\ 6 & 2 & 1\end{bmatrix} = \begin{bmatrix}52 & 21 & 9\\ 44 & 23 & 8\\ 5 & 4 & 1\end{bmatrix}$$

Exercise $4.E.95$

$$\begin{bmatrix}-3 & 1 & 5\\ 1 & 3 & 3\\ 3 & -3 & -3\end{bmatrix}\begin{bmatrix}2 & 2 & 1\\ 1 & 2 & 1\\ 4 & 1 & 1\end{bmatrix} = \begin{bmatrix}15 & 1 & 3\\ 17 & 11 & 7\\ -9 & -3 & -3\end{bmatrix}$$

Exercise $4.E.96$

$$\begin{bmatrix}3 & 1 & 1\\ 3 & 2 & 3\\ 3 & 3 & -1\end{bmatrix}\begin{bmatrix}6 & 2 & 1\\ 5 & 2 & 1\\ 6 & 1 & 1\end{bmatrix} = \begin{bmatrix}29 & 9 & 5\\ 46 & 13 & 8\\ 27 & 11 & 5\end{bmatrix}$$

Exercise $4.E.97$

$$\begin{bmatrix}5 & 3 & 2\\ 2 & 3 & 5\\ 5 & 5 & -2\end{bmatrix}\begin{bmatrix}11 & 4 & 1\\ 10 & 4 & 1\\ 12 & 3 & 1\end{bmatrix} = \begin{bmatrix}109 & 38 & 10\\ 112 & 35 & 10\\ 81 & 34 & 8\end{bmatrix}$$

Exercise $4.E.98$

Exercise $4.E.99$

Exercise $4.E.101$

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 $4.E.102$

$$\frac{1}{35}\begin{bmatrix}1 & 5 & 3\\ 5 & 25 & 15\\ 3 & 15 & 9\end{bmatrix}$$

Exercise $4.E.103$

$$\frac{1}{10}\begin{bmatrix}1 & 0 & 3\\ 0 & 0 & 0\\ 3 & 0 & 9\end{bmatrix}$$

Exercise $4.E.105$

$$C_B(\vec{x}) = \begin{bmatrix}2\\ 1\\ -1\end{bmatrix}.$$

Exercise $4.E.106$

$$M_{B_2B_1} = \begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix}$$

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).

5: LINEAR TRANSFORMATIONS

Note 5.1

Learn about linear transformations and their relationship to matrices.

In practice, one is often lead to ask questions about the geometry of a *transformation*: a function that takes an input and produces an output. This kind of question can be answered by linear algebra if the transformation can be expressed by a matrix.

5.1: Prelude to Linear Transformations and Matrix Algebra

5.2: Linear Transformations

5.3: Matrix Transformations

5.4: One-to-one and Onto Transformations

5.5: The Invertible Matrix Theorem

This page titled 5: Linear Transformations is shared under a GNU FDL license and was authored, remixed, and/or curated by Dan Margalit and Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- 3: Linear Transformations and Matrix Algebra by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

5.1: Prelude to Linear Transformations and Matrix Algebra

Example 5.1.1

Suppose you are building a robot arm with three joints that can move its hand around a plane, as in the following picture.

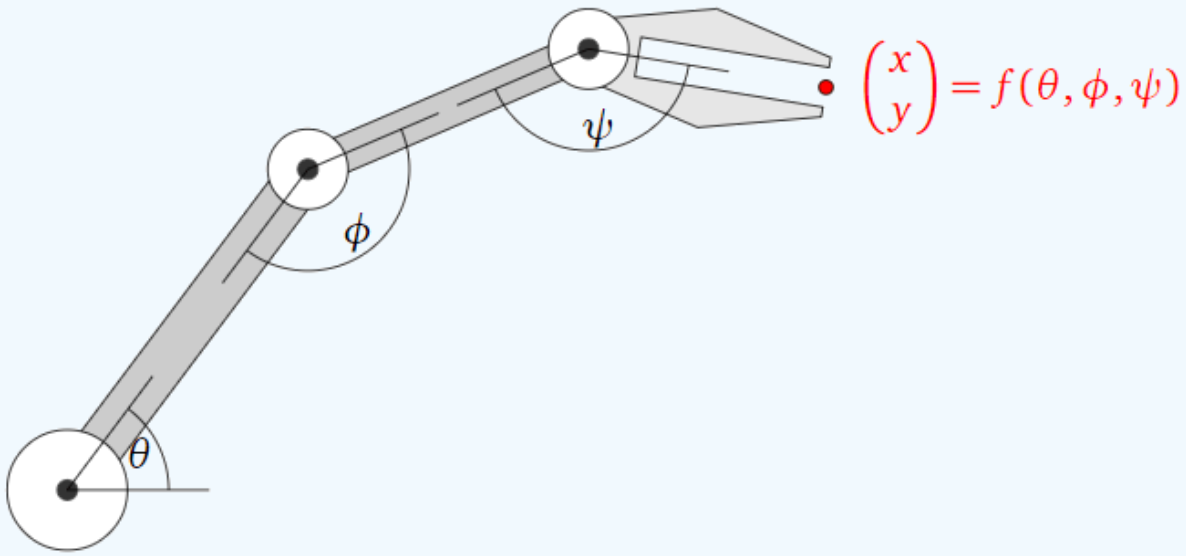


Figure 5.1.1

Jun 23, 2021, 12:13 PM

Define a transformation f as follows: $f(\theta, \phi, \psi)$ is the (x, y) position of the hand when the joints are rotated by angles θ, ϕ, ψ, respectively. The output of f tells you where the hand will be on the plane when the joints are set at the given input angles.

Unfortunately, this kind of function does not come from a matrix, so one cannot use linear algebra to answer questions about this function. In fact, these functions are rather complicated; their study is the subject of inverse kinematics.

In this chapter, we will be concerned with the relationship between matrices and transformations. In Section 3.1, we will consider the equation $b = Ax$ as a function with independent variable x and dependent variable b, and we draw pictures accordingly. We spend some time studying transformations in the abstract, and asking questions about a transformation, like whether it is one-to-one and/or onto (Section 3.2). In Section 3.3 we will answer the question: "when exactly can a transformation be expressed by a matrix?" We then present matrix multiplication as a special case of composition of transformations (Section 3.4). This leads to the study of *matrix algebra*: that is, to what extent one can do arithmetic with matrices in the place of numbers. With this in place, we learn to solve matrix equations by *dividing* by a matrix in Section 3.5.

This page titled 5.1: Prelude to Linear Transformations and Matrix Algebra is shared under a GNU FDL license and was authored, remixed, and/or curated by Dan Margalit and Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- 3.0: Prelude to Linear Transformations and Matrix Algebra by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

5.2: Linear Transformations

Objectives

1. Learn how to verify that a transformation is linear, or prove that a transformation is not linear.
2. Understand the relationship between linear transformations and matrix transformations.
3. *Recipe:* compute the matrix of a linear transformation.
4. *Theorem:* linear transformations and matrix transformations.
5. *Notation:* the ***standard coordinate vectors*** $e_1, e_2, \ldots$.
6. *Vocabulary words:* ***linear transformation***, ***standard matrix***, ***identity matrix***.

In Section 3.1, we studied the geometry of matrices by regarding them as functions, i.e., by considering the associated *matrix transformations*. We defined some vocabulary (domain, codomain, range), and asked a number of natural questions about a transformation. For a matrix transformation, these translate into questions about matrices, which we have many tools to answer.

In this section, we make a change in perspective. Suppose that we are given a *transformation* that we would like to study. If we can prove that our transformation is a matrix transformation, then we can use linear algebra to study it. This raises two important questions:

1. How can we tell if a transformation is a matrix transformation?
2. If our transformation is a matrix transformation, how do we find its matrix?

For example, we saw in Example 3.1.7 in Section 3.1 that the matrix transformation

$$T\colon \mathbb{R}^2 \to \mathbb{R}^2 \qquad T(x) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} x$$

is a counterclockwise rotation of the plane by 90°. However, we could have *defined* T in this way:

$$T\colon \mathbb{R}^2 \to \mathbb{R}^2 \qquad T(x) = \text{the counterclockwise rotation of } x \text{ by } 90^\circ.$$

Given this definition, it is not at all obvious that T is a matrix transformation, or what matrix it is associated to.

Linear Transformations: Definition

In this section, we introduce the class of transformations that come from matrices.

Definition 5.2.1: Linear Transformation

A ***linear transformation*** is a transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ satisfying

$$\begin{aligned} T(u+v) &= T(u) + T(v) \\ T(cu) &= cT(u) \end{aligned}$$

for all vectors u, v in $\mathbb{R}^n$ and all scalars c.

Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a matrix transformation: $T(x) = Ax$ for an $m \times n$ matrix A. By Proposition 2.3.1 in Section 2.3, we have

$$\begin{aligned} T(u+v) = A(u+v) &= Au + Av = T(u) + T(v) \\ T(cu) = A(cu) &= cAu = cT(u) \end{aligned}$$

for all vectors u, v in $\mathbb{R}^n$ and all scalars c. Since a matrix transformation satisfies the two defining properties, it is a linear transformation

We will see in the next Subsection The Matrix of a Linear Transformation that the opposite is true: every linear transformation is a matrix transformation; we just haven't computed its matrix yet.

Fact 5.2.1: Facts about Linear Transformations

Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then:

1. $T(0) = 0$.
2. For any vectors $v_1, v_2, \ldots, v_k$ in $\mathbb{R}^n$ and scalars $c_1, c_2, \ldots, c_k$, we have
$$T(c_1v_1 + c_2v_2 + \cdots + c_kv_k) = c_1T(v_1) + c_2T(v_2) + \cdots + c_kT(v_k).$$

Proof available at the end of this section.

In engineering, the second fact is called the *superposition principle*; it should remind you of the distributive property. For example, $T(cu + dv) = cT(u) + dT(v)$ for any vectors u, v and any scalars c, d. To restate the first fact:

Note 5.2.1

A linear transformation necessarily takes the zero vector to the zero vector.

Example 5.2.1: A non-linear transformation

Define $T: \mathbb{R} \to \mathbb{R}$ by $T(x) = x + 1$. Is T a linear transformation?

Solution

We have $T(0) = 0 + 1 = 1$. Since any linear transformation necessarily takes zero to zero by the above important note 5.2.1, we conclude that T is *not* linear (even though its graph is a line).

Note: in this case, it was not necessary to check explicitly that T does not satisfy both defining properties, Definition 5.2.1 : since $T(0) = 0$ is a consequence of these properties, at least one of them must not be satisfied. (In fact, this T satisfies neither.)

Example 5.2.2: Verifying linearity: dilation

Define $T: \mathbb{R}^2 \to \mathbb{R}^2$ by $T(x) = 1.5x$. Verify that T is linear.

Solution

We have to check the defining properties, Definition 5.2.1, for *all* vectors u, v and *all* scalars c. In other words, we have to treat $u, v,$ and c as *unknowns*. The only thing we are allowed to use is the definition of T.

$$\begin{aligned} T(u+v) &= 1.5(u+v) = 1.5u + 1.5v = T(u) + T(v) \\ T(cu) &= 1.5(cu) = c(1.5u) = cT(u). \end{aligned}$$

Since T satisfies both defining properties, T is linear.

Note: we know from Example 3.1.5 in Section 3.1 that T is a matrix transformation: in fact,

$$T(x) = \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix} x.$$

Since a matrix transformation is a linear transformation, this is another proof that T is linear.

Example 5.2.3: Verifying linearity: rotation

Define $T: \mathbb{R}^2 \to \mathbb{R}^2$ by

$$T(x) = \text{ the vector } x \text{ rotated counterclockwise by the angle } \theta.$$

Verify that T is linear.

Solution

Since T is defined geometrically, we give a geometric argument. For the first property, $T(u) + T(v)$ is the sum of the vectors obtained by rotating u and v by θ. On the other side of the equation, $T(u + v)$ is the vector obtained by rotating the sum of the vectors u and v. But it does not matter whether we sum or rotate first, as the following picture shows.

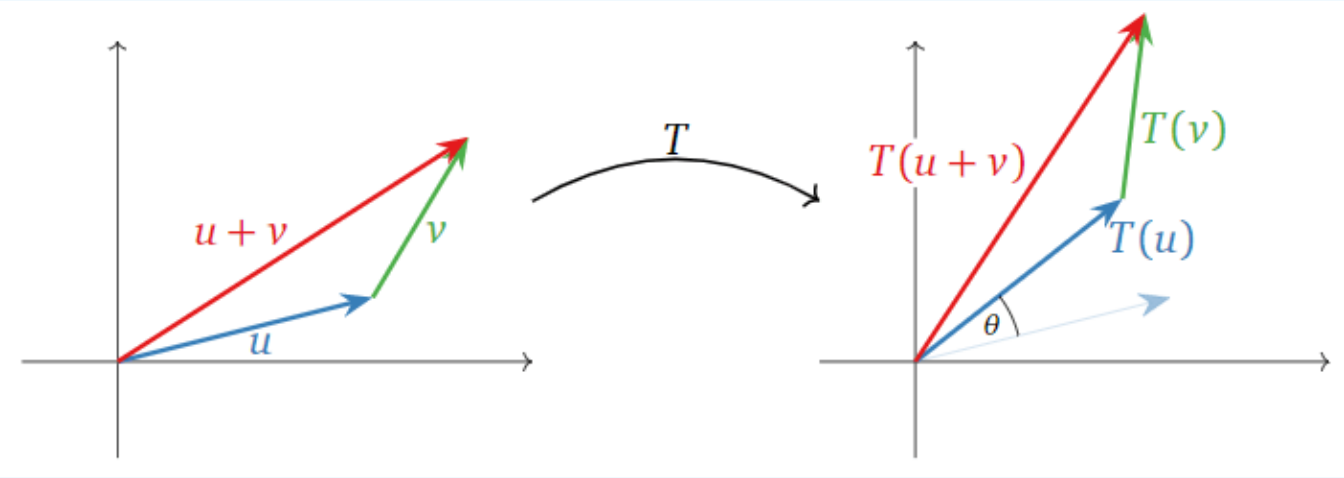


Figure 5.2.1

For the second property, $cT(u)$ is the vector obtained by rotating u by the angle θ, then changing its length by a factor of c (reversing direction of $c < 0$. On the other hand, $T(cu)$ first changes the length of c, then rotates. But it does not matter in which order we do these two operations.

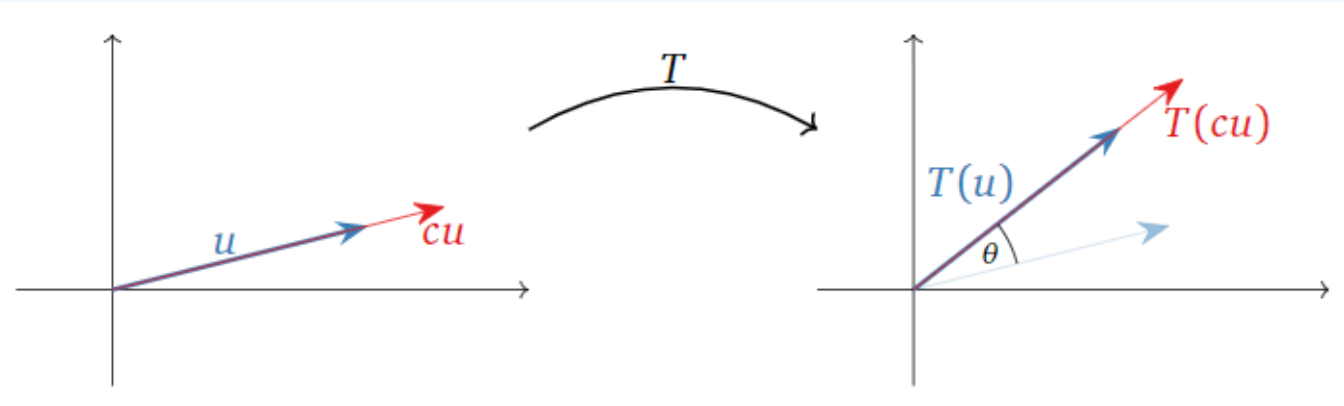


Figure 5.2.2

This verifies that T is a linear transformation. We will find its matrix in the next Subsection The Matrix of a Linear Transformation. Note however that it is not at all obvious that T can be expressed as multiplication by a matrix.

Example 5.2.4: A transformation defined by a formula

Define $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ by the formula

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3x - y \\ y \\ x \end{pmatrix}.$$

Verify that T is linear.

Solution

We have to check the defining properties, Definition 5.2.1, for *all* vectors u, v and *all* scalars c. In other words, we have to treat u, v, and c as *unknowns*; the only thing we are allowed to use is the definition of T. Since T is defined in terms of the coordinates of u, v, we need to give those names as well; say $u = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ and $v = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$. For the first property, we have

$$\begin{aligned} T\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) &= T\begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} = \begin{pmatrix} 3(x_1 + x_2) - (y_1 + y_2) \\ y_1 + y_2 \\ x_1 + x_2 \end{pmatrix} \\ &= \begin{pmatrix} (3x_1 - y_1) + (3x_2 - y_2) \\ y_1 + y_2 \\ x_1 + x_2 \end{pmatrix} \\ &= \begin{pmatrix} 3x_1 - y_1 \\ y_1 \\ x_1 \end{pmatrix} + \begin{pmatrix} 3x_2 - y_2 \\ y_2 \\ x_2 \end{pmatrix} = T\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + T\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}. \end{aligned}$$

For the second property,

$$\begin{aligned} T\left(c\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right) &= T\begin{pmatrix} cx_1 \\ cy_1 \end{pmatrix} = \begin{pmatrix} 3(cx_1) - (cy_1) \\ cy_1 \\ cx_1 \end{pmatrix} \\ &= \begin{pmatrix} c(3x_1 - y_1) \\ cy_1 \\ cx_1 \end{pmatrix} = c\begin{pmatrix} 3x_1 - y_1 \\ y_1 \\ x_1 \end{pmatrix} = cT\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}. \end{aligned} \tag{5.2.1}$$

Since T satisfies the defining properties, Definition 5.2.1, T is a linear transformation.

Note: we will see in this Example 5.2.9 below that

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Hence T is in fact a matrix transformation.

One can show that, if a transformation is defined by formulas in the coordinates as in the above example, then the transformation is linear if and only if each coordinate is a linear expression in the variables with no constant term.

Example 5.2.5: A translation

Define $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ by

$$T(x) = x + \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

This kind of transformation is called a ***translation***. As in a previous Example 5.2.1, this T is not linear, because $T(0)$ is not the zero vector.

Example 5.2.6: More non-linear transformations

Verify that the following transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$ are not linear:

$$T_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} |x| \\ y \end{pmatrix} \quad T_2 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} xy \\ y \end{pmatrix} \quad T_3 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x+1 \\ x-2y \end{pmatrix}.$$

Solution

In order to verify that a transformation T is *not* linear, we have to show that T does not satisfy *at least one* of the two defining properties, Definition 5.2.1. For the first, the negation of the statement "$T(u+v) = T(u) + T(v)$ for all vectors u, v" is "there exists at least one pair of vectors u, v such that $T(u+v) \neq T(u) + T(v)$." In other words, it suffices to find *one example* of a pair of vectors u, v such that $T(u+v) \neq T(u) + T(v)$. Likewise, for the second, the negation of the statement "$T(cu) = cT(u)$ for all vectors u and all scalars c" is "there exists some vector u and some scalar c such that $T(cu) \neq cT(u)$." In other words, it suffices to find *one* vector u and *one* scalar c such that $T(cu) \neq cT(u)$.

For the first transformation, we note that

$$T_1\left(-\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = T_1 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} |-1| \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

but that

$$-T_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = -\begin{pmatrix} |1| \\ 0 \end{pmatrix} = -\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

Therefore, this transformation does not satisfy the second property.

For the second transformation, we note that

$$T_2\left(2\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) = T_2 \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \cdot 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$$

but that

$$2T_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2\begin{pmatrix} 1 \cdot 1 \\ 1 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}.$$

Therefore, this transformation does not satisfy the second property.

For the third transformation, we observe that

$$T_3 \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2(0)+1 \\ 0-2(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Since T_3 does not take the zero vector to the zero vector, it cannot be linear.

When deciding whether a transformation T is linear, generally the first thing to do is to check whether $T(0) = 0$; if not, T is automatically not linear. Note however that the non-linear transformations T_1 and T_2 of the above example do take the zero vector

to the zero vector.

Challenge

Find an example of a transformation that satisfies the first property of linearity, Definition 5.2.1, but not the second.

The Standard Coordinate Vectors

In the next subsection, we will present the relationship between linear transformations and matrix transformations. Before doing so, we need the following important notation.

Note 5.2.2: Standard Coordinate Vectors

The ***standard coordinate vectors in*** $\mathbb{R}^n$ are the n vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad \cdots, \quad e_{n-1} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

The ith entry of e_i is equal to 1, and the other entries are zero.

From now on, for the rest of the book, we will use the symbols $e_1, e_2, \ldots$ to denote the standard coordinate vectors.

There is an ambiguity in this notation: one has to know from context that e_1 is meant to have n entries. That is, the vectors

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

may both be denoted e_1, depending on whether we are discussing vectors in $\mathbb{R}^2$ or in $\mathbb{R}^3$.

The standard coordinate vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ are pictured below.

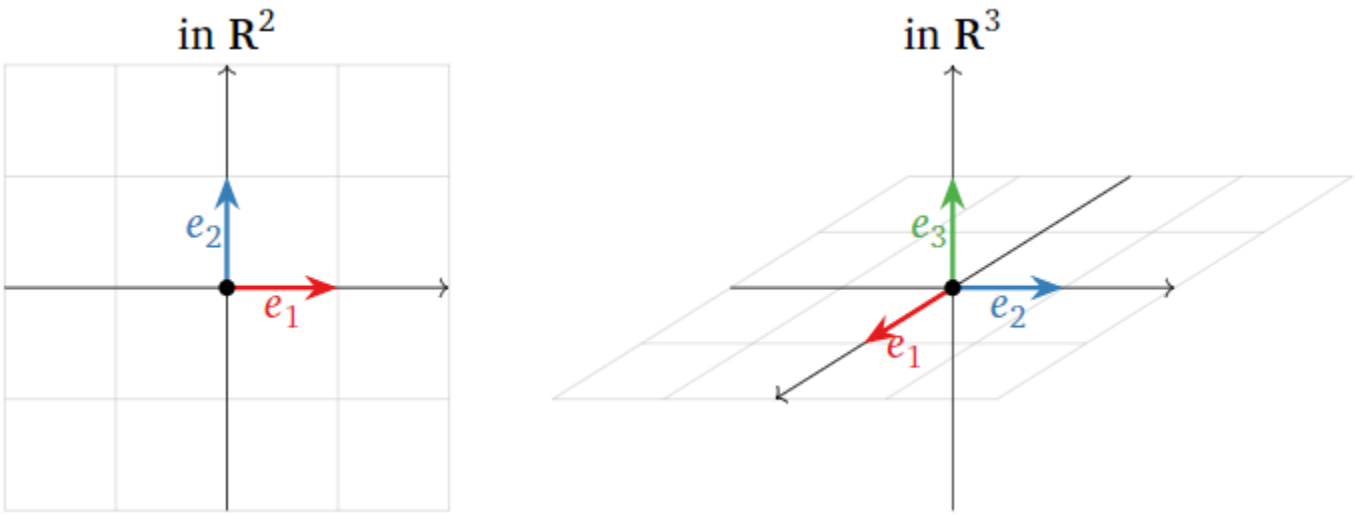


Figure 5.2.3

These are the vectors of length 1 that point in the positive directions of each of the axes.

Fact 5.2.2: Multiplying a matrix by the standard coordinate vectors

If A is an $m \times n$ matrix with columns $v_1, v_2, \ldots, v_m$, then $Ae_i = v_i$ for each $i = 1, 2, \ldots, n$:

$$\begin{pmatrix} | & | & & | \\ v_1 & v_2 & & v_n \\ | & | & & | \end{pmatrix} e_i = v_i.$$

In other words, multiplying a matrix by e_i simply selects its ith column.

For example,

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 7 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ 8 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 01 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix}.$$

Definition 5.2.2: Identity Matrix

The $n \times n$ ***identity matrix*** is the matrix I_n whose columns are the n standard coordinate vectors in $\mathbb{R}^n$:

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

We will see in this Example 5.2.10 below that the identity matrix is the matrix of the identity transformation, Definition 3.1.2.

The Matrix of a Linear Transformation

Now we can prove that every linear transformation is a matrix transformation, and we will show how to compute the matrix.

Theorem 5.2.1: The Matrix of a Linear Transformation

Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Let A be the $m \times n$ matrix

$$A = \begin{pmatrix} | & | & & | \\ T(e_1) & T(e_2) & \cdots & T(e_n) \\ | & | & & | \end{pmatrix}.$$

Then T is the matrix transformation associated with A: that is, $T(x) = Ax$.

Proof available at the end of this section.

The matrix A in the above theorem is called the ***standard matrix*** for T. The columns of A are the vectors obtained by evaluating

T on the n standard coordinate vectors in $\mathbb{R}^n$. To summarize part of the theorem:

Note 5.2.3

Matrix transformations are the same as linear transformations.

Note 5.2.4: Dictionary

Linear transformations are the same as matrix transformations, which come from matrices. The correspondence can be summarized in the following dictionary.

$$\begin{array}{cc} T:\mathbb{R}^n\to\mathbb{R}^m & \\ \text{Linear transformation} & \longrightarrow m\times n \text{ matrix } A=\begin{pmatrix} | & | & & | \\ T(e_1) & T(e_2) & \cdots & T(e_n) \\ | & | & & | \end{pmatrix} \\ T:\mathbb{R}^n\to\mathbb{R}^m & \\ T(x)=Ax & \longleftarrow m\times n \text{ matrix } A \end{array} \tag{5.2.2}$$

Example 5.2.7: The matrix of a dilation

Define $T\colon\mathbb{R}^2\to\mathbb{R}^2$ by $T(x)=1.5x$. Find the standard matrix A for T.

Solution

The columns of A are obtained by evaluating T on the standard coordinate vectors e_1, e_2.

$$\left.\begin{aligned} T(e_1)=1.5e_1=\begin{pmatrix}1.5\\0\end{pmatrix} \\ T(e_2)=1.5e_2=\begin{pmatrix}0\\1.5\end{pmatrix}\end{aligned}\right\} \implies A=\begin{pmatrix}1.5 & 0\\0 & 1.5\end{pmatrix}.$$

This is the matrix we started with in Example 3.1.5 in Section 3.1.

Example 5.2.8: The matrix of a rotation

Define $T\colon\mathbb{R}^2\to\mathbb{R}^2$ by

$$T(x)= \text{ the vector } x \text{ rotated counterclockwise by the angle } \theta.$$

Find the standard matrix for T.

Solution

The columns of A are obtained by evaluating T on the standard coordinate vectors e_1, e_2. In order to compute the entries of $T(e_1)$ and $T(e_2)$, we have to do some trigonometry.

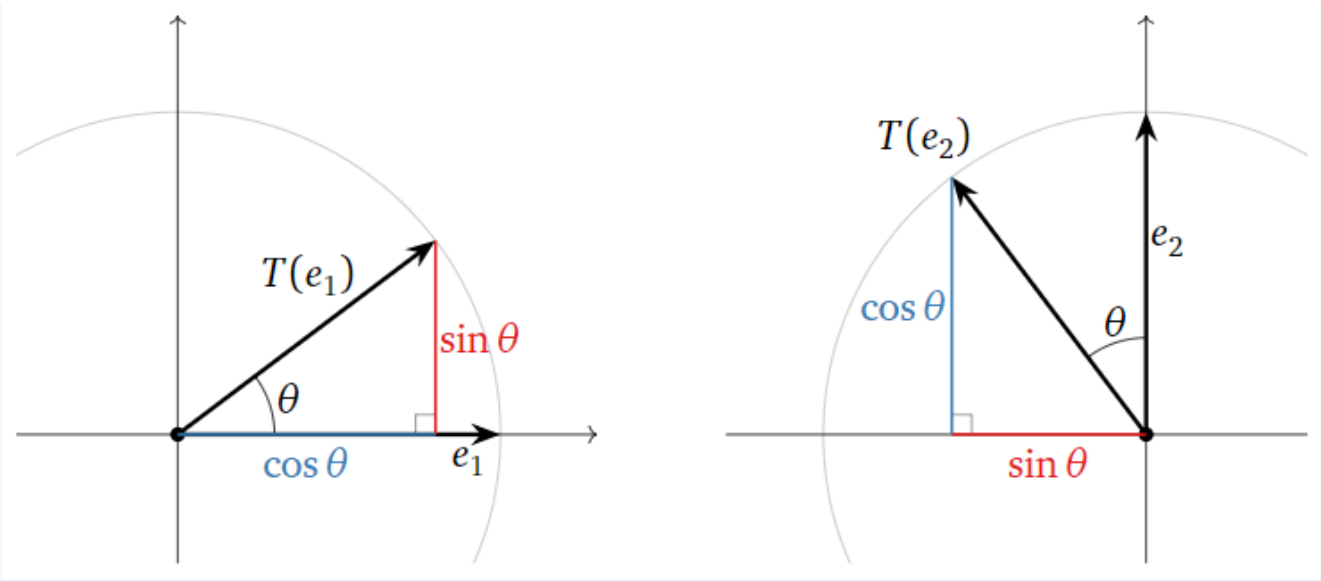


Figure 5.2.4

We see from the picture that

$$\left.\begin{aligned} T(e_1) &= \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \\ T(e_2) &= \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} \end{aligned}\right\} \implies A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

We saw in the above example that the matrix for counterclockwise rotation of the plane by an angle of θ is

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Example 5.2.9: A transformation defined by a formula

Define $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ by the formula

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3x - y \\ y \\ x \end{pmatrix}.$$

Find the standard matrix for T.

Solution

We substitute the standard coordinate vectors into the formula defining T:

$$\left.\begin{aligned} T(e_1) = T\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3(1)-0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} \\ T(e_2) = T\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 3(0)-1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \end{aligned}\right\} \Longrightarrow A = \begin{pmatrix} 3 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Example 5.2.10: A transformation defined in steps

Let $T: \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation that reflects over the xy-plane and then projects onto the yz-plane. What is the standard matrix for T?

Solution

This transformation is described geometrically, in two steps. To find the columns of A, we need to follow the standard coordinate vectors through each of these steps.

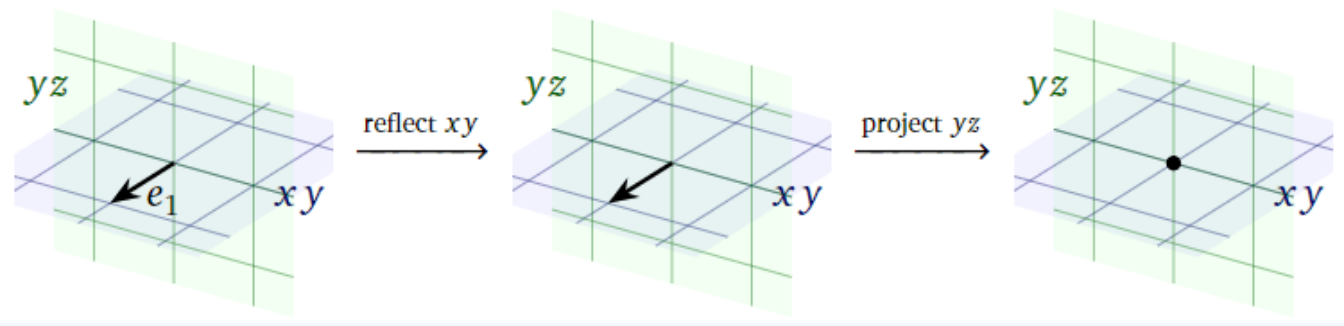


Figure 5.2.5

Since e_1 lies on the xy-plane, reflecting over the xy-plane does not move e_1. Since e_1 is perpendicular to the yz-plane, projecting e_1 onto the yz-plane sends it to zero. Therefore,

$$T(e_1) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

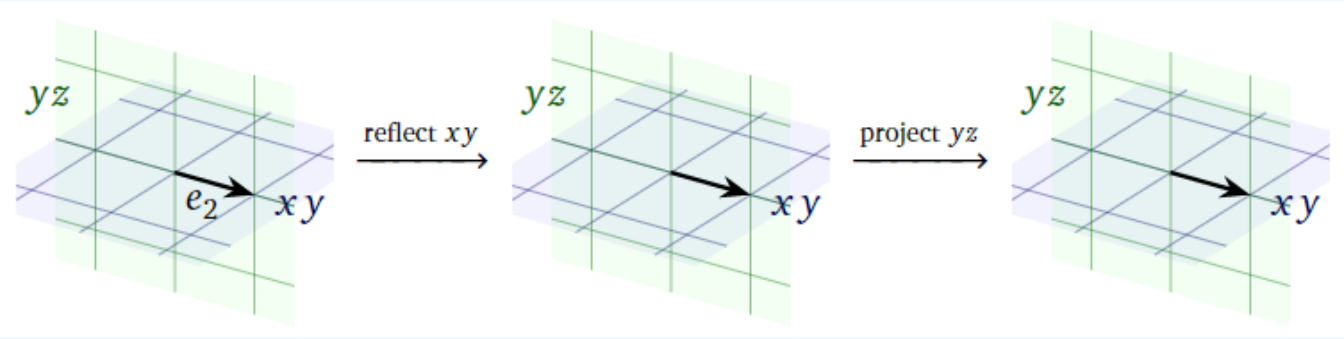


Figure 5.2.6

Since e_2 lies on the xy-plane, reflecting over the xy-plane does not move e_2. Since e_2 lies on the yz-plane, projecting onto the yz-plane does not move e_2 either. Therefore,

$$T(e_2) = e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

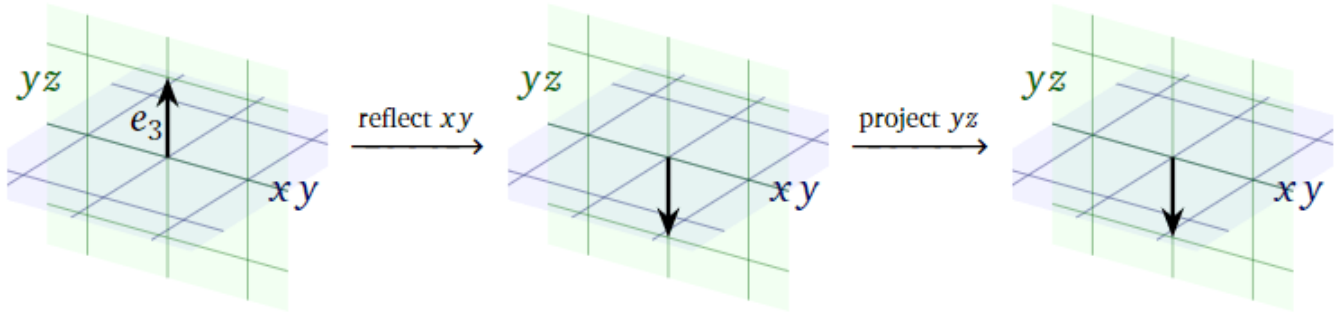


Figure 5.2.7

Since e_3 is perpendicular to the xy-plane, reflecting over the xy-plane takes e_3 to its negative. Since $-e_3$ lies on the yz-plane, projecting onto the yz-plane does not move it. Therefore,

$$T(e_3) = -e_3 = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}.$$

Now we have computed all three columns of A:

$$\left.\begin{aligned} T(e_1) &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ T(e_2) &= \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\ T(e_1) &= \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \end{aligned}\right\} \Longrightarrow A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

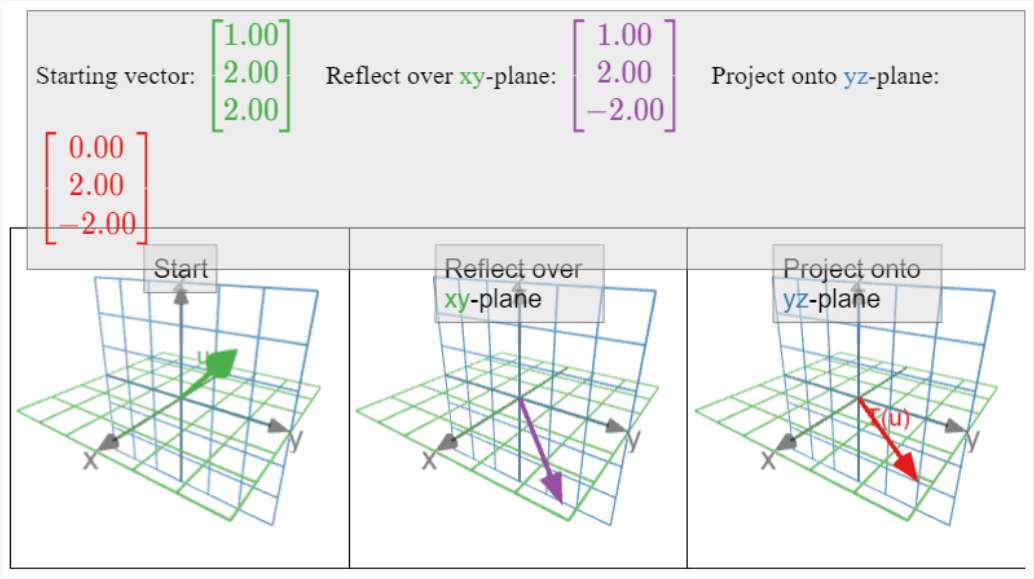


Figure 5.2.8: Illustration of a transformation defined in steps. Click and drag the vector on the left.

Recall from Definition 3.1.2 in Section 3.1 that the *identity transformation* is the transformation $\mathrm{Id}_{\mathbb{R}^n}\colon \mathbb{R}^n \to \mathbb{R}^n$ defined by $\mathrm{Id}_{\mathbb{R}^n}(x) = x$ for every vector x.

Example 5.2.11: The standard matrix of the identity transformation

Verify that the identity transformation $\mathrm{Id}_{\mathbb{R}^n}\colon \mathbb{R}^n \to \mathbb{R}^n$ is linear, and compute its standard matrix.

Solution

We verify the two defining properties, Definition 5.2.1, of linear transformations. Let u, v be vectors in $\mathbb{R}^n$. Then

$$\mathrm{Id}_{\mathbb{R}^n}(u+v) = u+v = \mathrm{Id}_{\mathbb{R}^n}(u) + \mathrm{Id}_{\mathbb{R}^n}(v).$$

If c is a scalar, then

$$\mathrm{Id}_{\mathbb{R}^n}(cu) = cu = c\mathrm{Id}_{\mathbb{R}^n}(u).$$

Since $\mathrm{Id}_{\mathbb{R}^n}$ satisfies the two defining properties, it is a linear transformation.

Now that we know that $\mathrm{Id}_{\mathbb{R}^n}$ is linear, it makes sense to compute its standard matrix. For each standard coordinate vector e_i, we have $\mathrm{Id}_{R^n}(e_i) = e_i$. In other words, the columns of the standard matrix of $\mathrm{Id}_{\mathbb{R}^n}$ are the standard coordinate vectors, so the standard matrix is the identity matrix

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

We computed in Example 5.2.11 that the matrix of the identity transform is the identity matrix: for every x in $\mathbb{R}^n$,

$$x = \mathrm{Id}_{\mathbb{R}^n}(x) = I_n x.$$

Therefore, $I_n x = x$ for all vectors x: the product of the identity matrix and a vector is the same vector.

Answers

Fact 5.2.1: Facts about Linear Transformations

1. Since $0 = -0$, we have
$$T(0) = T(-0) = -T(0)$$
by the second defining property, Definition 5.2.1. The only vector w such that $w = -w$ is the zero vector.
2. Let us suppose for simplicity that $k = 2$. Then
$$\begin{aligned} T(c_1v_1 + c_2v_2) &= T(c_1v_1) + T(c_2v_2) && \text{first property} \\ &= c_1T(v_1) + c_2T(v_2) && \text{second property.} \end{aligned}$$

Theorem 5.2.1: The Matrix of a Linear Transformation

We suppose for simplicity that T is a transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$. Let A be the matrix given in the statement of the theorem. Then

$$\begin{aligned} T\begin{pmatrix} x \\ y \\ z \end{pmatrix}; &= T\left(x\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + z\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right) \\ &= T(xe_1 + ye_2 + ze_3) \\ &= xT(e_1) + yT(e_2) + zT(e_3) \\ &= \begin{pmatrix} | & | & | \\ T(e_1) & T(e_2) & T(e_3) \\ | & | & | \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= A\begin{pmatrix} x \\ y \\ z \end{pmatrix}. \end{aligned}$$

This page titled 5.2: Linear Transformations is shared under a GNU FDL license and was authored, remixed, and/or curated by Dan Margalit and Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- 3.3: Linear Transformations by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

5.3: Matrix Transformations

Objectives

1. Learn to view a matrix geometrically as a function.
2. Learn examples of matrix transformations: reflection, dilation, rotation, shear, projection.
3. Understand the vocabulary surrounding transformations: domain, codomain, range.
4. Understand the domain, codomain, and range of a matrix transformation.
5. *Pictures:* common matrix transformations.
6. *Vocabulary words:* ***transformation*** / ***function***, ***domain***, ***codomain***, ***range***, ***identity transformation***, ***matrix transformation***.

In this section we learn to understand matrices geometrically as *functions*, or *transformations*. We briefly discuss transformations in general, then specialize to matrix transformations, which are transformations that come from matrices.

Matrices as Functions

Informally, a *function* is a rule that accepts inputs and produces outputs. For instance, $f(x) = x^2$ is a function that accepts one number x as its input, and outputs the square of that number: $f(2) = 4$. In this subsection, we interpret matrices as functions.

Let A be a matrix with m rows and n columns. Consider the matrix equation $b = Ax$ (we write it this way instead of $Ax = b$ to remind the reader of the notation $y = f(x)$). If we vary x, then b will also vary; in this way, we think of A as a function with independent variable x and dependent variable b.

- The independent variable (the input) is x, which is a vector in $\mathbb{R}^n$.
- The dependent variable (the output) is b, which is a vector in $\mathbb{R}^m$.

The set of all possible output vectors are the vectors b such that $Ax = b$ has some solution; this is the same as the column space of A by Note 2.3.6 in Section 2.3.

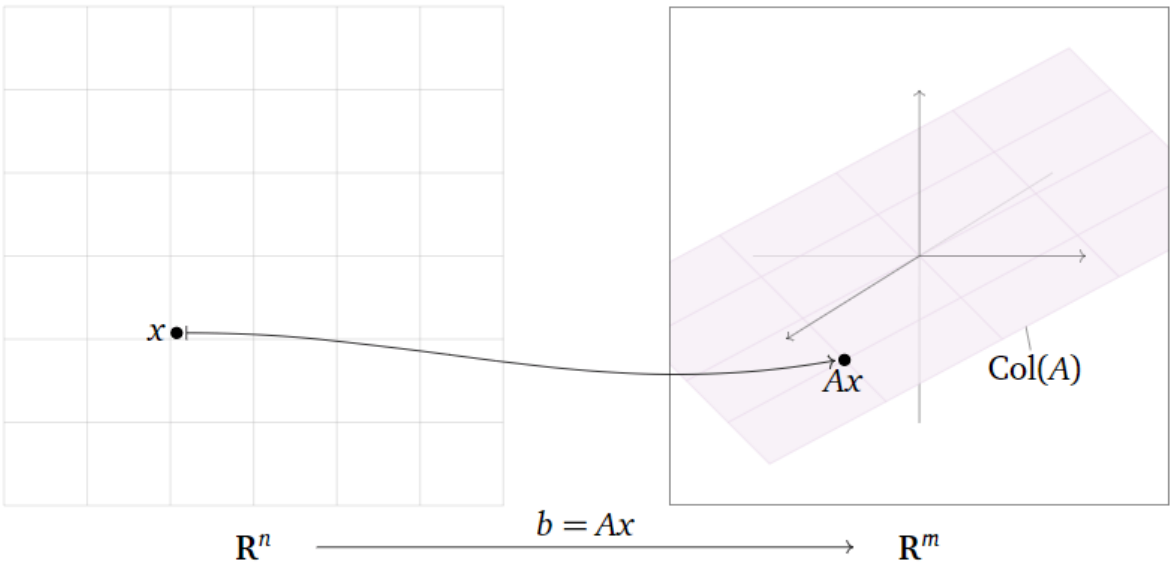


Figure 5.3.1

Example 5.3.1: Interactive: A 2×3 matrix

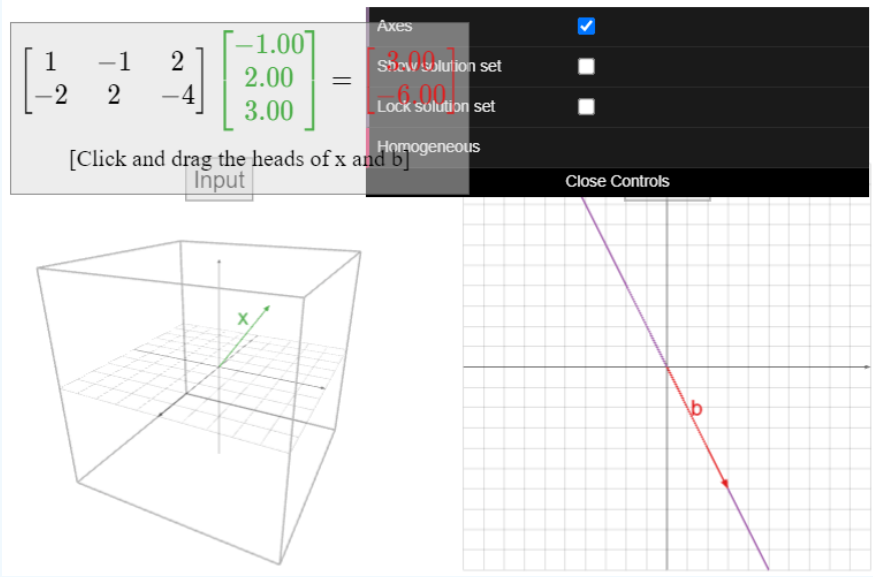


Figure 5.3.2: A picture of a 2×3 matrix, regarded as a function. The input vector is x, which is a vector in $\mathbb{R}^3$, and the output vector is $b = Ax$, which is a vector in $\mathbb{R}^2$. The violet line on the right is the column space; as you vary x, the output b is constrained to lie on this line.

Example 5.3.2: Interactive: A 3×2 matrix

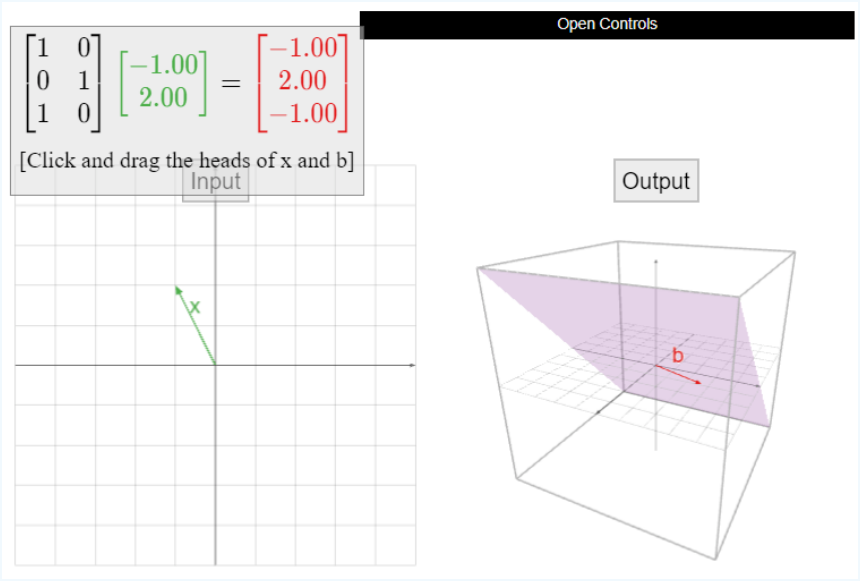


Figure 5.3.3: A picture of a 3×2 matrix, regarded as a function. The input vector is x, which is a vector in $\mathbb{R}^2$, and the output vector is $b = Ax$, which is a vector in $\mathbb{R}^3$. The violet plane on the right is the column space; as you vary x, the output b is constrained to lie on this plane.

Example 5.3.3: Projection onto the xy-plane

Let

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

Describe the function $b = Ax$ geometrically.

Solution

In the equation $Ax = b$, the input vector x and the output vector b are both in $\mathbb{R}^3$. First we multiply A by a vector to see what it does:

$$A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}.$$

Multiplication by A simply sets the z-coordinate equal to zero: it *projects vertically onto the xy-plane*.

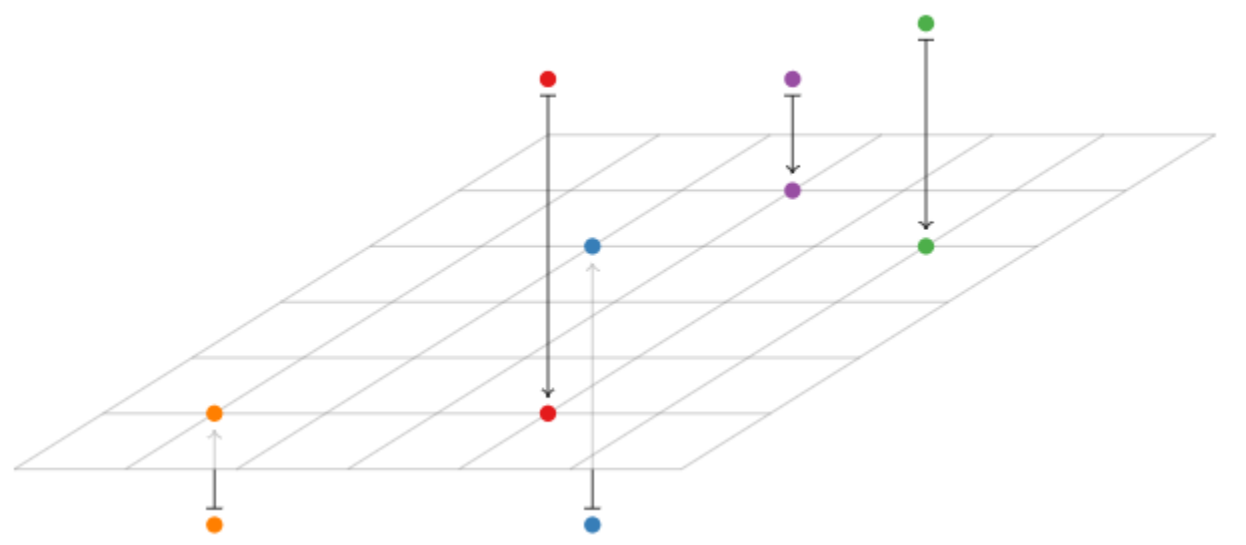


Figure 5.3.4

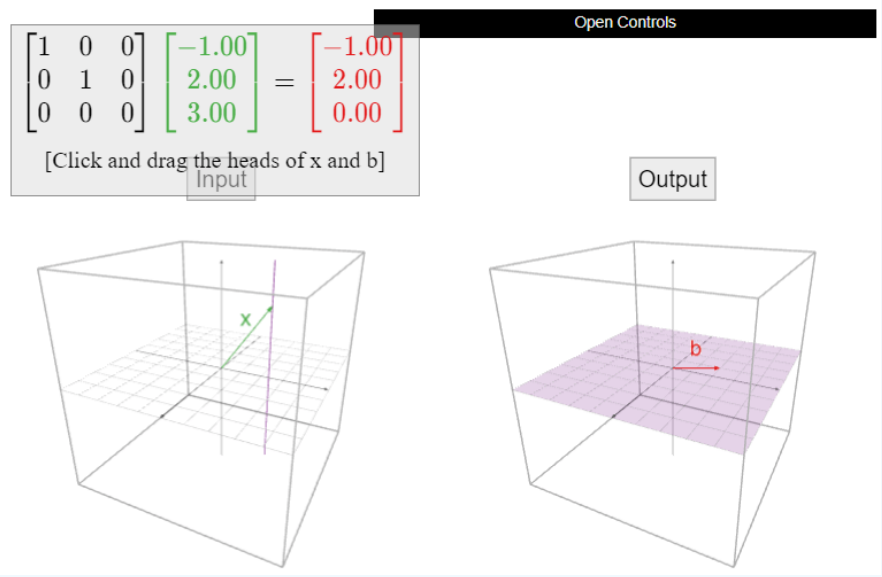


Figure 5.3.5: Multiplication by the matrix A projects a vector onto the xy-plane. Move the input vector x to see how the output vector b changes.

Example 5.3.4: Reflection

Let

$$A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Describe the function $b = Ax$ geometrically.

Solution

In the equation $Ax = b$, the input vector x and the output vector b are both in $\mathbb{R}^2$. First we multiply A by a vector to see what it does:

$$A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x \\ y \end{pmatrix}.$$

Multiplication by A negates the x-coordinate: it *reflects over the y-axis*.

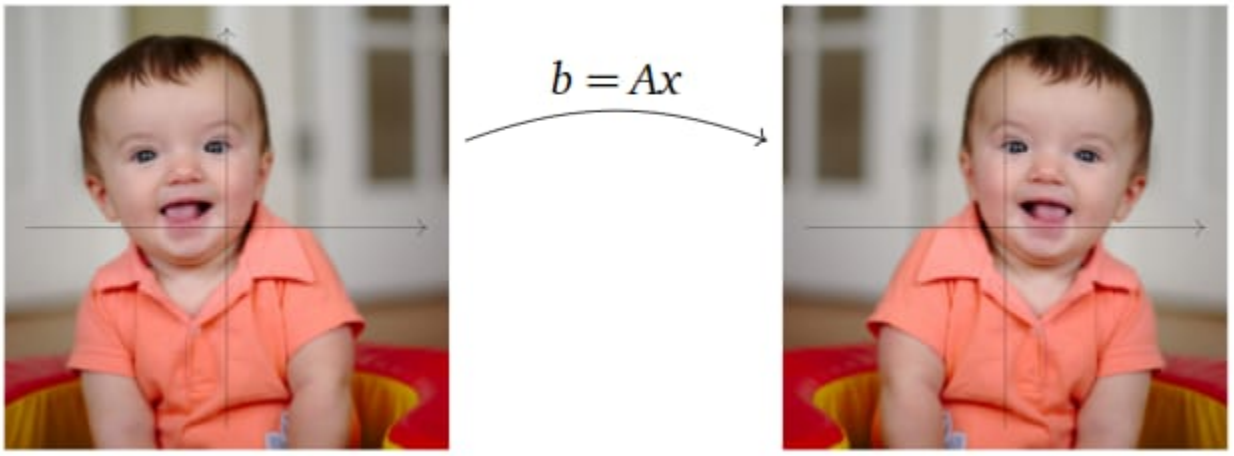


Figure 5.3.6

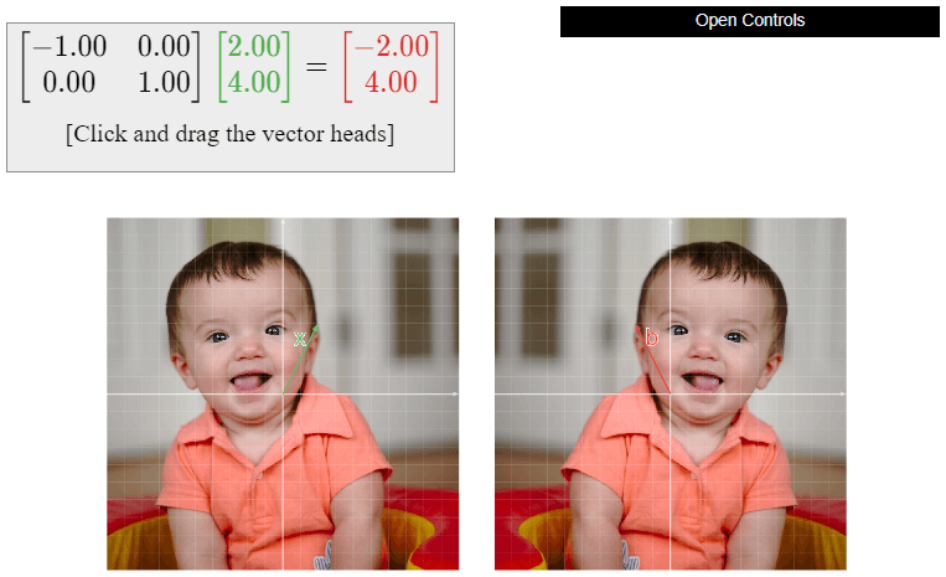


Figure 5.3.7: Multiplication by the matrix A reflects over the y-axis. Move the input vector x to see how the output vector b changes.

Example 5.3.5: Dilation

Let

$$A = \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix}.$$

Describe the function $b = Ax$ geometrically.

Solution

In the equation $Ax = b$, the input vector x and the output vector b are both in $\mathbb{R}^2$. First we multiply A by a vector to see what it does:

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.5x \\ 1.5y \end{pmatrix} = 1.5 \begin{pmatrix} x \\ y \end{pmatrix}.$$

Multiplication by A is the same as scalar multiplication by 1.5: it *scale*s or *dilate*s the plane by a factor of 1.5.

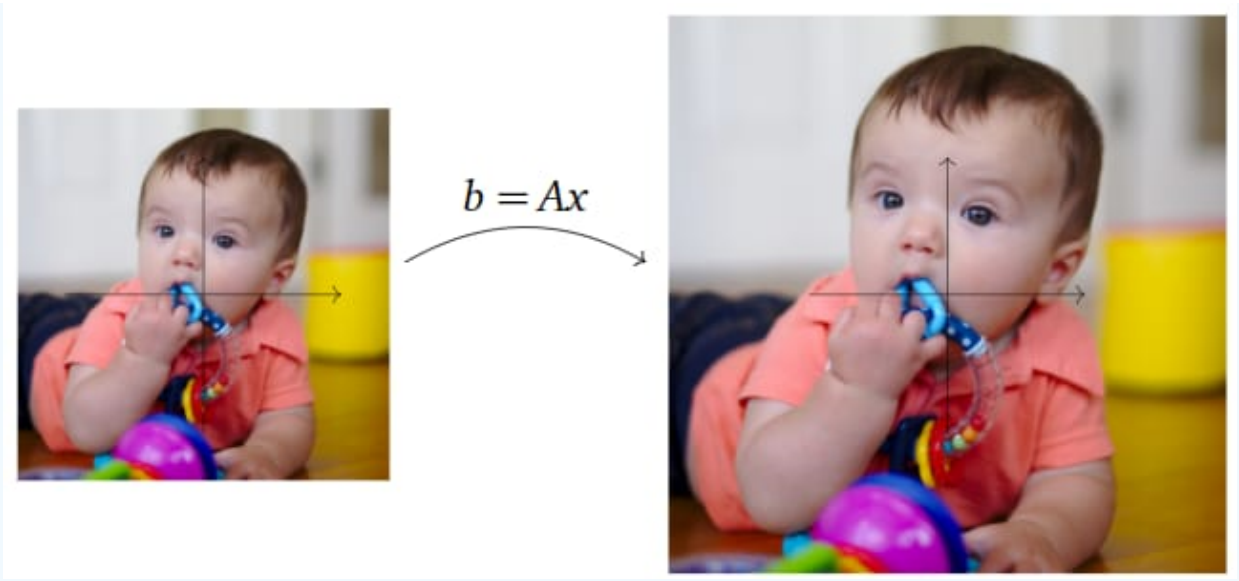


Figure 5.3.8

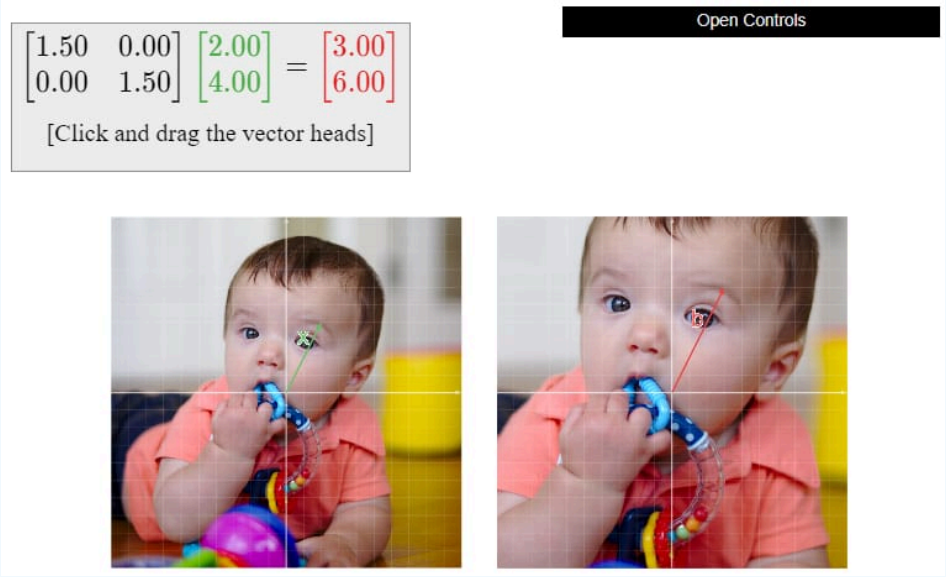


Figure 5.3.9: Multiplication by the matrix A dilates the plane by a factor of 1.5. Move the input vector x to see how the output vector b changes.

Example 5.3.6: Identity

Let

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

Describe the function $b = Ax$ geometrically.

Solution

In the equation $Ax = b$, the input vector x and the output vector b are both in $\mathbb{R}^2$. First we multiply A by a vector to see what it does:

$$A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Multiplication by A does not change the input vector at all: it is the *identity transformation which does nothing*.

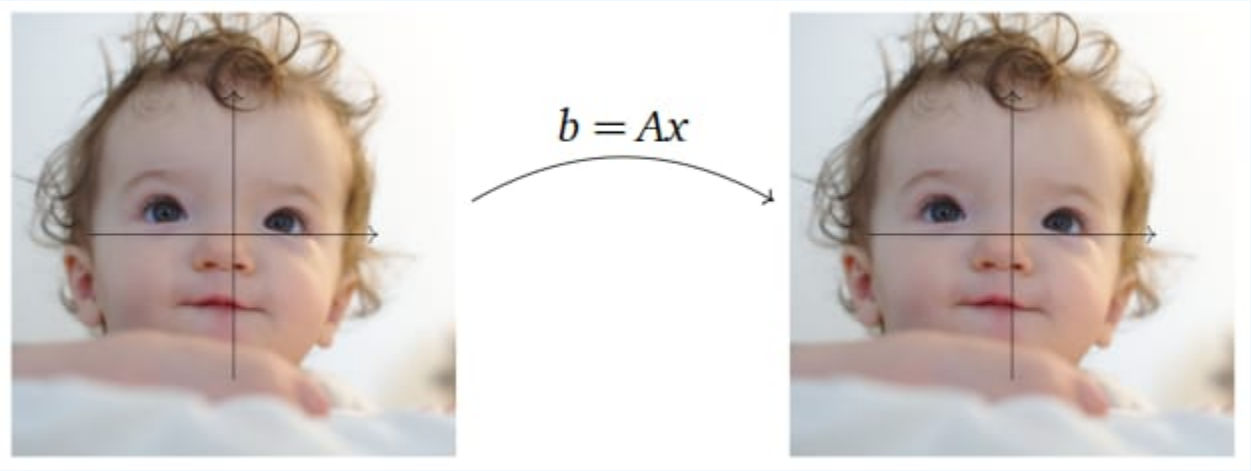


Figure 5.3.10

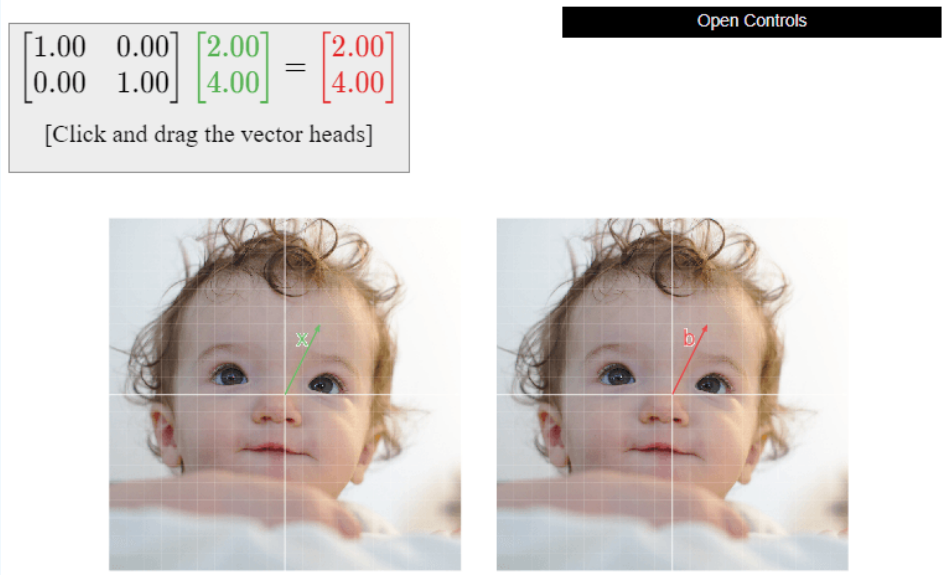


Figure 5.3.11: Multiplication by the matrix A does not move the vector x: that is, $b = Ax = x$. Move the input vector x to see how the output vector b changes.

Example 5.3.7: Rotation

Let

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Describe the function $b = Ax$ geometrically.

Solution

In the equation $Ax = b$, the input vector x and the output vector b are both in $\mathbb{R}^2$. First we multiply A by a vector to see what it does:

$$A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}.$$

We substitute a few test points in order to understand the geometry of the transformation:

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

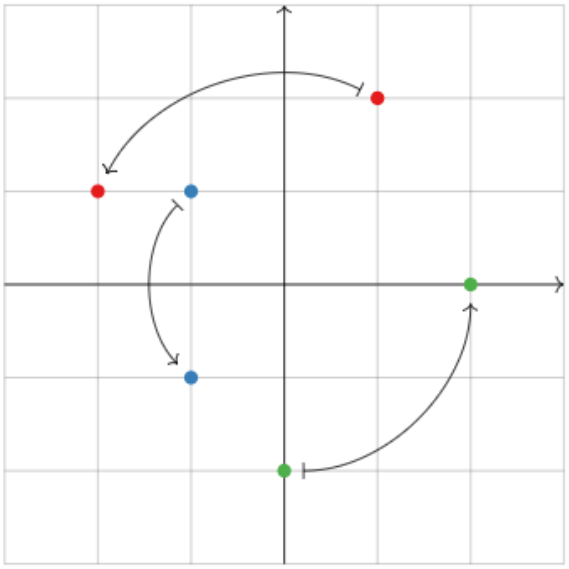


Figure 5.3.12

Multiplication by A is *counterclockwise rotation by* $90°$.

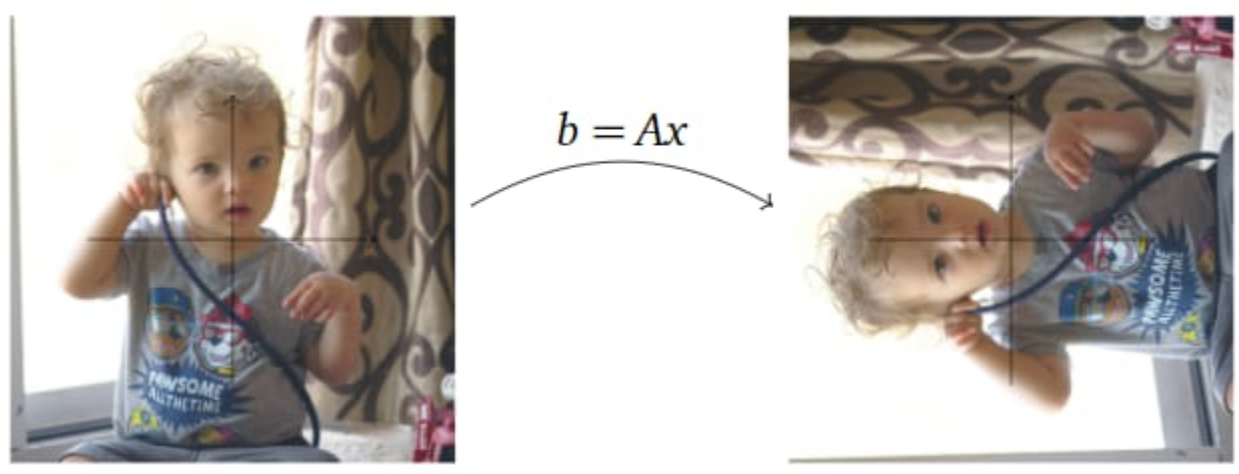


Figure 5.3.13

$$\begin{bmatrix} 0.00 & -1.00 \\ 1.00 & 0.00 \end{bmatrix} \begin{bmatrix} 2.00 \\ 4.00 \end{bmatrix} = \begin{bmatrix} -4.00 \\ 2.00 \end{bmatrix}$$

[Click and drag the vector heads]

Open Controls

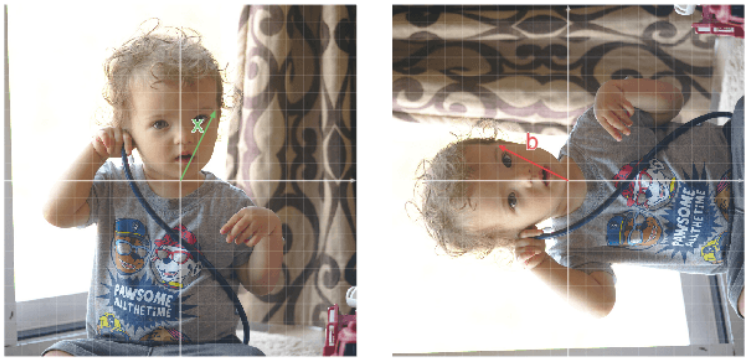


Figure 5.3.14: Multiplication by the matrix A rotates the vector x counterclockwise by 90°. Move the input vector x to see how the output vector b changes.

Example 5.3.8: Shear

Let

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

Describe the function $b = Ax$ geometrically.

Solution

In the equation $Ax = b$, the input vector x and the output vector b are both in $\mathbb{R}^2$. First we multiply A by a vector to see what it does:

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ y \end{pmatrix}.$$

Multiplication by A adds the y-coordinate to the x-coordinate; this is called a *shear in the x-direction*.

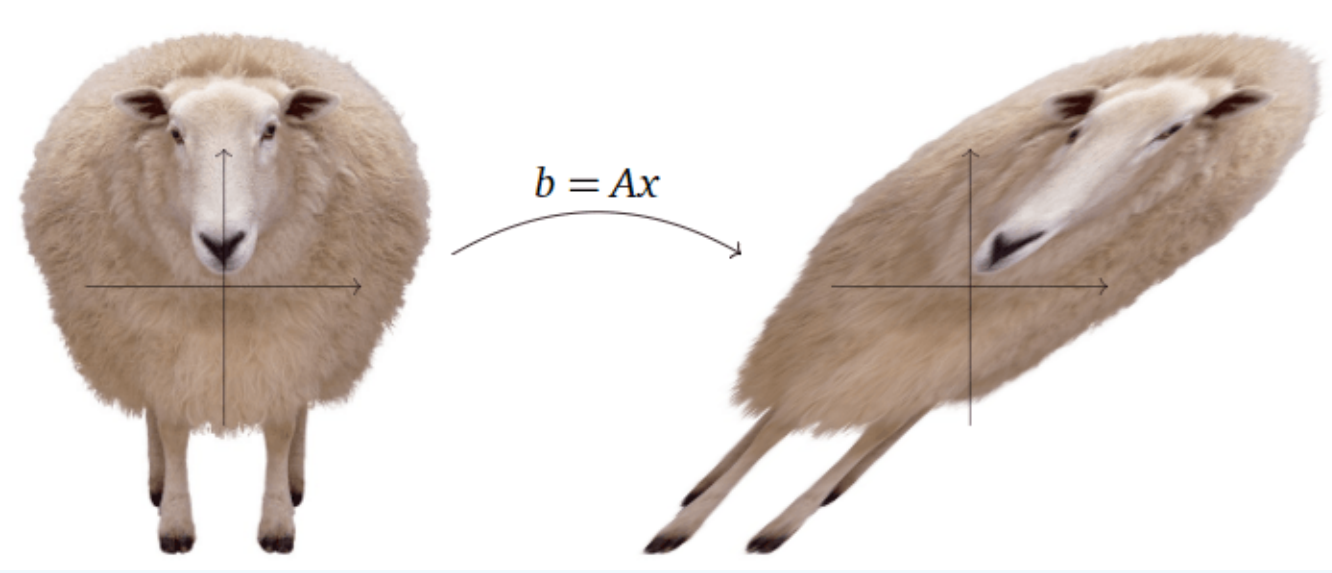


Figure 5.3.15

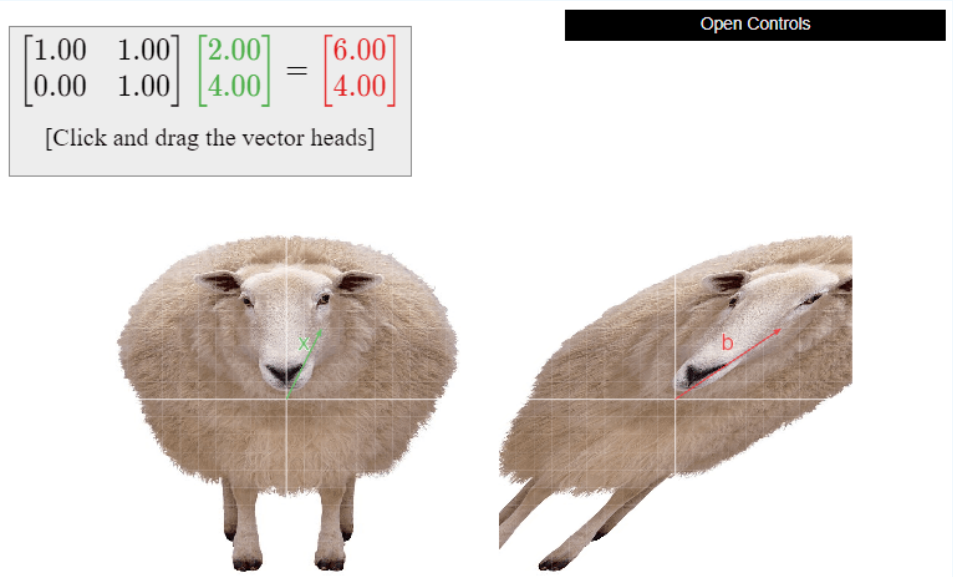


Figure 5.3.16: Multiplication by the matrix A adds the y-coordinate to the x-coordinate. Move the input vector x to see how the output vector b changes.

Transformations

At this point it is convenient to fix our ideas and terminology regarding functions, which we will call *transformations* in this book. This allows us to systematize our discussion of matrices as functions.

Definition 5.3.1: Transformation

A ***transformation*** from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule T that assigns to each vector x in $\mathbb{R}^n$ a vector $T(x)$ in $\mathbb{R}^m$.

- $\mathbb{R}^n$ is called the ***domain*** of T.
- $\mathbb{R}^m$ is called the ***codomain*** of T.

- For x in $\mathbb{R}^n$, the vector $T(x)$ in $\mathbb{R}^m$ is the ***image*** of x under T.
- The set of all images $\{T(x) \mid x \text{ in } \mathbb{R}^n\}$ is the ***range*** of T.

The notation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ means "T is a transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$."

It may help to think of T as a "machine" that takes x as an input, and gives you $T(x)$ as the output.

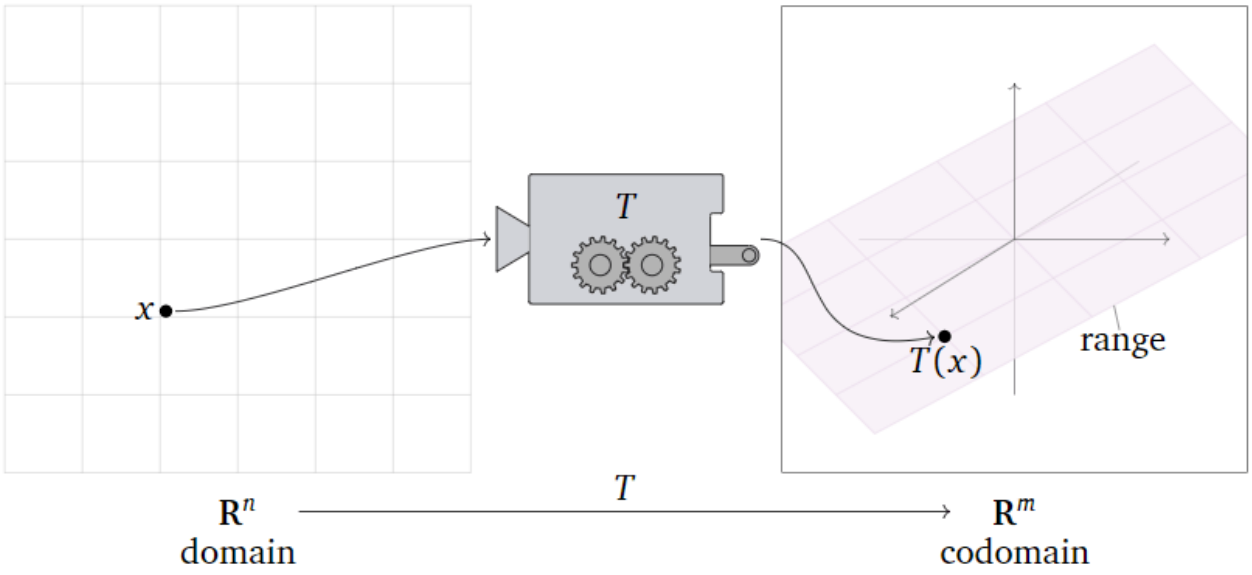


Figure 5.3.17

The points of the domain $\mathbb{R}^n$ are the *inputs* of T: this simply means that it makes sense to evaluate T on vectors with n entries, i.e., lists of n numbers. Likewise, the points of the codomain $\mathbb{R}^m$ are the *outputs* of T: this means that the result of evaluating T is always a vector with m entries.

The *range* of T is the set of all vectors in the codomain that actually arise as outputs of the function T, for some input. In other words, the range is all vectors b in the codomain such that $T(x) = b$ has a solution x in the domain.

Example 5.3.9: A Function of one variable

Most of the functions you may have seen previously have domain and codomain equal to $\mathbb{R} = \mathbb{R}^1$. For example,

$$\sin : \mathbb{R} \to \mathbb{R} \quad \sin(x) = \begin{pmatrix} \text{the length of the opposite} \\ \text{edge over the hypotenuse of} \\ \text{a right triangle with angle } x \\ \text{in radians} \end{pmatrix}.$$

Notice that we have defined $\sin$ by a rule: a function is defined by specifying what the output of the function is for any possible input.

You may be used to thinking of such functions in terms of their graphs:

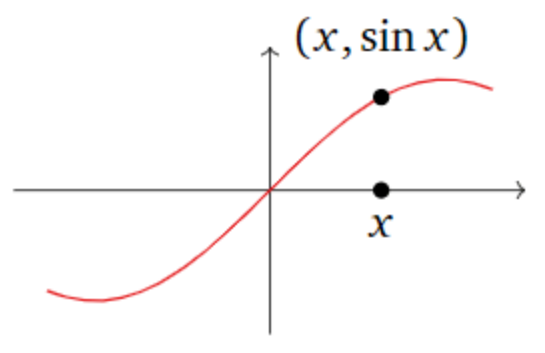


Figure 5.3.18

In this case, the horizontal axis is the domain, and the vertical axis is the codomain. This is useful when the domain and codomain are $\mathbb{R}$, but it is hard to do when, for instance, the domain is $\mathbb{R}^2$ and the codomain is $\mathbb{R}^3$. The graph of such a function is a subset of $\mathbb{R}^5$, which is difficult to visualize. For this reason, we will rarely graph a transformation.

Note that the *range* of $\sin$ is the interval $[-1, 1]$: this is the set of all possible outputs of the $\sin$ function.

Example 5.3.10: Functions of several variables

Here is an example of a function from $\mathbb{R}^2$ to $\mathbb{R}^3$:

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ \cos(y) \\ y-x^2 \end{pmatrix}.$$

The inputs of f each have two entries, and the outputs have three entries. In this case, we have defined f by a formula, so we evaluate f by substituting values for the variables:

$$f\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2+3 \\ \cos(3) \\ 3-2^2 \end{pmatrix} = \begin{pmatrix} 5 \\ \cos(3) \\ -1 \end{pmatrix}.$$

Here is an example of a function from $\mathbb{R}^3$ to $\mathbb{R}^3$:

$$f(v) = \begin{pmatrix} \text{the counterclockwise rotation} \\ \text{of } v \text{ by and angle of } 42^\circ \text{ about} \\ \text{the } z\text{-axis} \end{pmatrix}.$$

In other words, f takes a vector with three entries, then rotates it; hence the ouput of f also has three entries. In this case, we have defined f by a geometric rule.

Definition 5.3.2: Identity Transformation

The ***identity transformation*** $\mathrm{Id}_{\mathbb{R}^n}\colon \mathbb{R}^n \to \mathbb{R}^n$ is the transformation defined by the rule

$$\mathrm{Id}_{\mathbb{R}^n}(x) = x \qquad \text{for all } x \text{ in } \mathbb{R}^n.$$

In other words, the identity transformation does not move its input vector: the output is the same as the input. Its domain and

codomain are both $\mathbb{R}^n$, and its range is $\mathbb{R}^n$ as well, since every vector in $\mathbb{R}^n$ is the output of itself.

Example 5.3.11: A real-word transformation: robotics

The definition of *transformation* and its associated vocabulary may seem quite abstract, but transformations are extremely common in real life. Here is an example from the fields of robotics and computer graphics.

Suppose you are building a robot arm with three joints that can move its hand around a plane, as in the following picture.

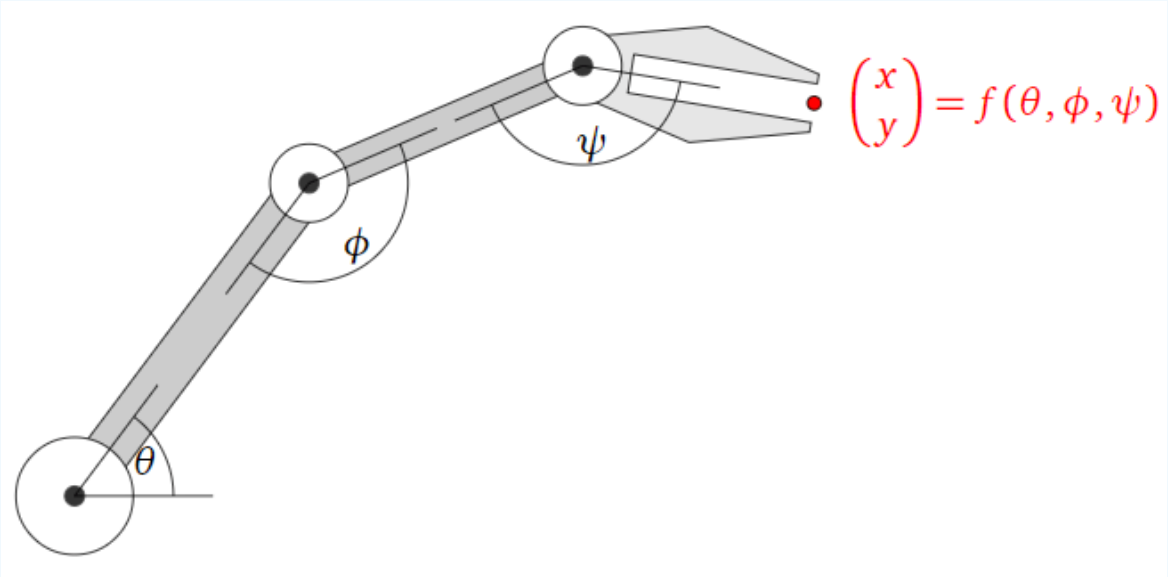


Figure 5.3.19

Define a transformation $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ as follows: $f(\theta, \phi, \psi)$ is the (x, y) position of the hand when the joints are rotated by angles θ, ϕ, ψ, respectively. Evaluating f tells you where the hand will be on the plane when the joints are set at the given angles.

It is relatively straightforward to find a formula for $f(\theta, \phi, \psi)$ using some basic trigonometry. If you want the robot to fetch your coffee cup, however, you have to *find* the angles θ, ϕ, ψ that will put the hand at the position of your beverage. It is not at all obvious how to do this, and it is not even clear if the answer is unique! You can ask yourself: "which positions on the table can my robot arm reach?" or "what is the arm's range of motion?" This is the same as asking: "what is the range of f?"

Unfortunately, this kind of function does not come from a matrix, so one cannot use linear algebra to answer these kinds of questions. In fact, these functions are rather complicated; their study is the subject of inverse kinematics.

Matrix Transformations

Now we specialize the general notions and vocabulary from the previous Subsection Transformations to the functions defined by matrices that we considered in the first Subsection Matrices as Functions.

Definition 5.3.3: Matrix Transformation

Let A be an $m \times n$ matrix. The ***matrix transformation*** associated to A is the transformation

$$T\colon \mathbb{R}^n \to \mathbb{R}^m \quad \text{defined by} \quad T(x) = Ax.$$

This is the transformation that takes a vector x in $\mathbb{R}^n$ to the vector Ax in $\mathbb{R}^m$.

If A has n columns, then it only makes sense to multiply A by vectors with n entries. This is why the domain of $T(x) = Ax$ is $\mathbb{R}^n$. If A has n rows, then Ax has m entries for any vector x in $\mathbb{R}^n$; this is why the codomain of $T(x) = Ax$ is $\mathbb{R}^m$.

The definition of a matrix transformation T tells us how to evaluate T on any given vector: we multiply the input vector by a matrix. For instance, let

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

and let $T(x) = Ax$ be the associated matrix transformation. Then

$$T\begin{pmatrix} -1 \\ -2 \\ -3 \end{pmatrix} = A\begin{pmatrix} -1 \\ -2 \\ -3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} -1 \\ -2 \\ -3 \end{pmatrix} = \begin{pmatrix} -14 \\ -32 \end{pmatrix}.$$

Suppose that A has columns $v_1, v_2, \ldots, v_n$. If we multiply A by a general vector x, we get

$$Ax = \begin{pmatrix} | & | & & | \\ v_1 & v & 2 & \cdots & v_n \\ | & | & & | \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1v_1 + x_2v_2 + \cdots + x_nv_n.$$

This is just a general linear combination of $v_1, v_2, \ldots, v_n$. Therefore, the outputs of $T(x) = Ax$ are exactly the linear combinations of the columns of A: the *range* of T is the column space of A. See Note 2.3.6 in Section 2.3.

Note 5.3.1

Let A be an $m \times n$ matrix, and let $T(x) = Ax$ be the associated matrix transformation.

- The *domain* of T is $\mathbb{R}^n$, where n is the number of *columns* of A.
- The *codomain* of T is $\mathbb{R}^m$, where m is the number of *rows* of A.
- The *range* of T is the *column space* of A.

Example 5.3.12: Interactive: A 2×3 matrix: reprise

Let

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

and define $T(x) = Ax$. The domain of T is $\mathbb{R}^3$, and the codomain is $\mathbb{R}^2$. The range of T is the column space; since all three columns are collinear, the range is a line in $\mathbb{R}^2$.

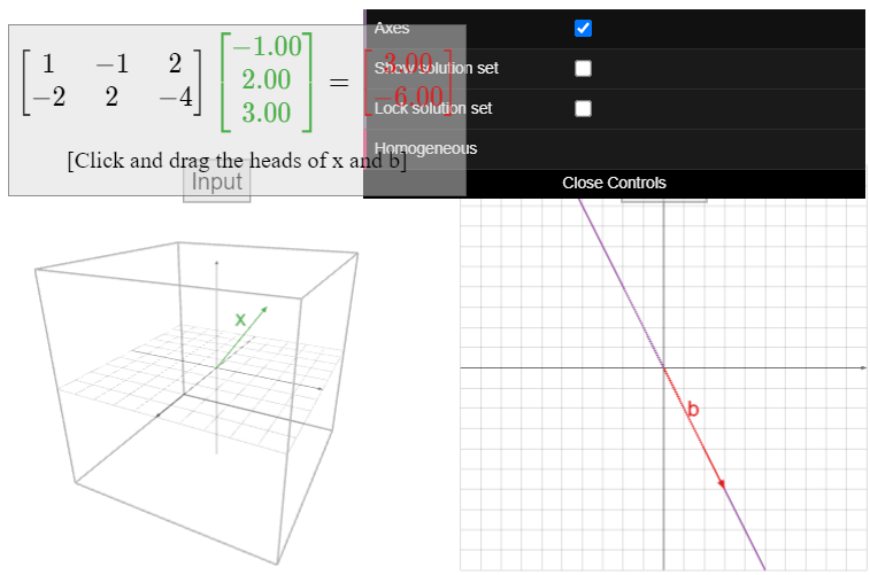


Figure 5.3.20: A picture of the matrix transformation T. The input vector is x, which is a vector in $\mathbb{R}^3$, and the output vector is $b = T(x) = Ax$, which is a vector in $\mathbb{R}^2$. The violet line on the right is the range of T; as you vary x, the output b is constrained to lie on this line.

Example 5.3.13: Interactive: A 3×2 matrix: reprise

Let

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

and define $T(x) = Ax$. The domain of T is $\mathbb{R}^2$, and the codomain is $\mathbb{R}^3$. The range of T is the column space; since A has two columns which are not collinear, the range is a plane in $\mathbb{R}^3$.

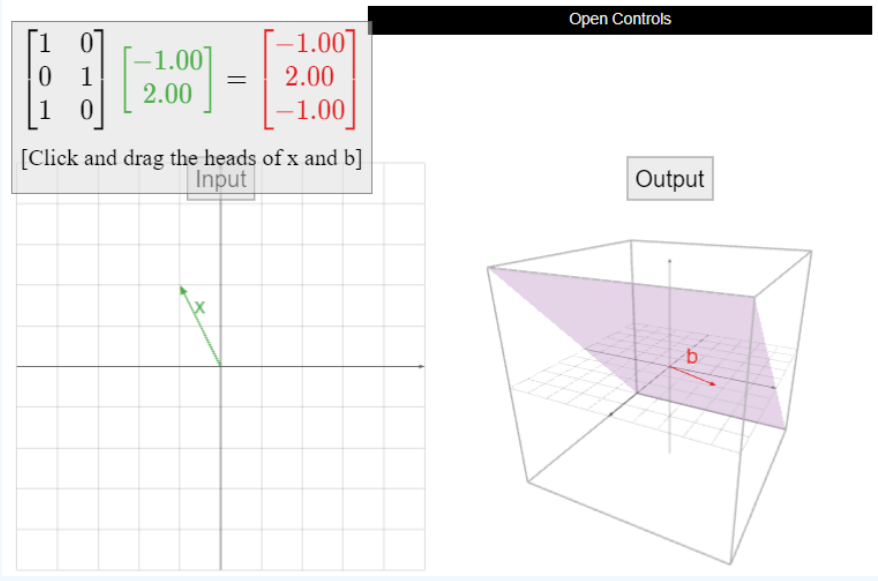


Figure 5.3.21: A picture of the matrix transformation T. The input vector is x, which is a vector in $\mathbb{R}^2$, and the output vector is $b = T(x) = Ax$, which is a vector in $\mathbb{R}^3$. The violet plane on the right is the range of T; as you vary x, the output b is constrained to lie on this plane.

Example 5.3.14: Projection onto the xy-plane: reprise

Let

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

and let $T(x) = Ax$. What are the domain, the codomain, and the range of T?

Solution

Geometrically, the transformation T projects a vector directly "down" onto the xy-plane in $\mathbb{R}^3$.

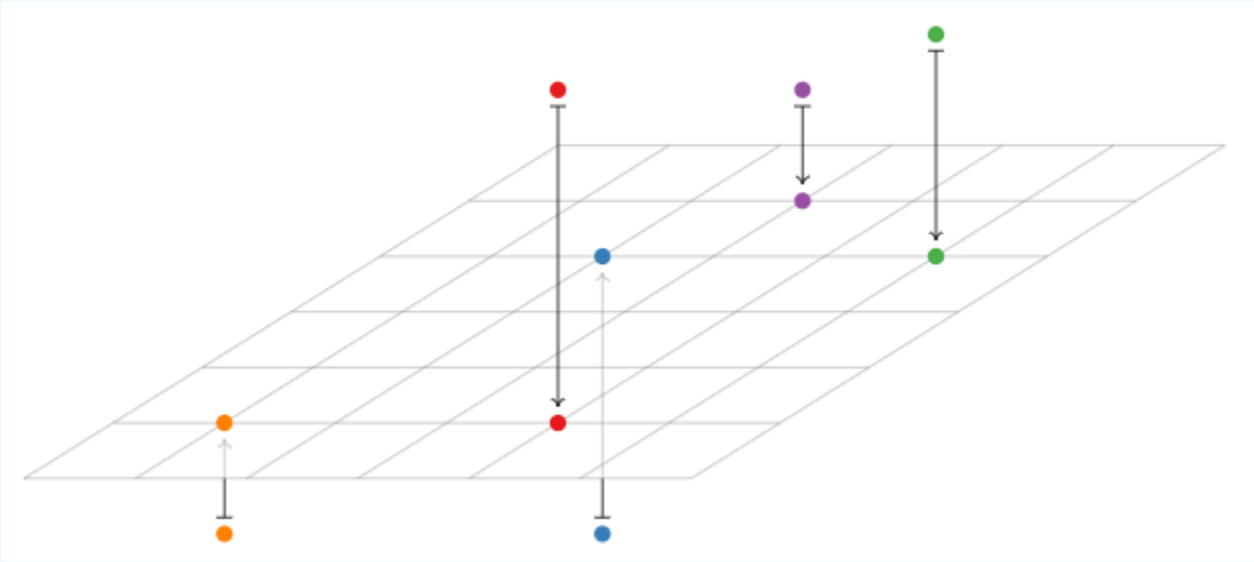


Figure 5.3.22

The inputs and outputs have three entries, so the domain and codomain are both $\mathbb{R}^3$. The possible outputs all lie on the xy-plane, and every point on the xy-plane is an output of T (with itself as the input), so the range of T is the xy-plane.

Be careful not to confuse the codomain with the range here. The range is a plane, but it is a plane *in* $\mathbb{R}^3$, so the codomain is still $\mathbb{R}^3$. The outputs of T all have three entries; the last entry is simply always zero.

In the case of an $n \times n$ square matrix, the domain and codomain of $T(x) = Ax$ are both $\mathbb{R}^n$. In this situation, one can regard T as *operating on* $\mathbb{R}^n$: it moves the vectors around in the same space.

Example 5.3.15: Matrix transformations of $\mathbb{R}^2$

In the first Subsection Matrices as Functions we discussed the transformations defined by several 2×2 matrices, namely:

$$\text{Reflection:} \quad A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$
$$\text{Dilation:} \quad A = \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix}$$
$$\text{Identity:} \quad A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$
$$\text{Rotation:} \quad A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$
$$\text{Shear:} \quad A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

In each case, the associated matrix transformation $T(x) = Ax$ has domain and codomain equal to $\mathbb{R}^2$. The range is also $\mathbb{R}^2$, as can be seen geometrically (what is the input for a given output?), or using the fact that the columns of A are not collinear (so they form a basis for $\mathbb{R}^2$).

Example 5.3.16: Questions about a [matrix] transformation

Let

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

and let $T(x) = Ax$, so $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ is a matrix transformation.

1. Evaluate $T(u)$ for $u = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$.
2. Let
$$b = \begin{pmatrix} 7 \\ 5 \\ 7 \end{pmatrix}.$$
Find a vector v in $\mathbb{R}^2$ such that $T(v) = b$. Is there more than one?
3. Does there exist a vector w in $\mathbb{R}^3$ such that there is more than one v in $\mathbb{R}^2$ with $T(v) = w$?
4. Find a vector w in $\mathbb{R}^3$ which is not in the range of T.

Note: all of the above questions are intrinsic to the transformation T: they make sense to ask whether or not T is a matrix transformation. See the next Example 5.3.17. As T is in fact a matrix transformation, all of these questions will translate into questions about the corresponding matrix A.

Solution

1. We evaluate $T(u)$ by substituting the definition of T in terms of matrix multiplication:
$$T\begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 7 \\ 4 \\ 7 \end{pmatrix}.$$
2. We want to find a vector v such that $b = T(v) = Av$. In other words, we want to solve the matrix equation $Av = b$. We form an augmented matrix and row reduce:
$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} v = \begin{pmatrix} 7 \\ 5 \\ 7 \end{pmatrix} \xrightarrow{\text{augmented matrix}} \begin{pmatrix} 1 & 1 & 7 \\ 0 & 1 & 5 \\ 1 & 1 & 7 \end{pmatrix} \xrightarrow{\text{row reduce}} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 5 \\ 0 & 0 & 0 \end{pmatrix}.$$

This gives $x = 2$ and $y = 5$, so that there is a unique vector

$$v = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$$

such that $T(v) = b$.

3. *Translation:* is there any vector w in $\mathbb{R}^3$ such that the solution set of $Av = w$ has more than one vector in it? The solution set of $Ax = w$, if non-empty, is a translate of the solution set of $Av = b$ above, which has one vector in it. See key observation 2.4.3 in Section 2.4. It follows that the solution set of $Av = w$ can have at most one vector.
4. *Translation:* find a vector w such that the matrix equation $Av = w$ is not consistent. Notice that if we take

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

then the matrix equation $Av = w$ translates into the system of equations

$$\begin{cases} x + y = 1 \\ \quad\; y = 2 \\ x + y = 3, \end{cases}$$

which is clearly inconsistent.

Example 5.3.17: Questions about a [non-matrix] transformation

Define a transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ by the formula

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \ln(x) \\ \cos(y) \\ \ln(x) \end{pmatrix}.$$

1. Evaluate $T(u)$ for $u = \begin{pmatrix} 1 \\ \pi \end{pmatrix}$.
2. Let

$$b = \begin{pmatrix} 7 \\ 1 \\ 7 \end{pmatrix}.$$

Find a vector v in $\mathbb{R}^2$ such that $T(v) = b$. Is there more than one?

3. Does there exist a vector w in $\mathbb{R}^3$ such that there is more than one v in $\mathbb{R}^2$ with $T(v) = w$?
4. Find a vector w in $\mathbb{R}^3$ which is not in the range of T.

Note: we asked (almost) the exact same questions about a matrix transformation in the previous Example 5.3.16. The point of this example is to illustrate the fact that the questions make sense for a transformation that has no hope of coming from a matrix. In this case, these questions do not translate into questions about a matrix; they have to be answered in some other way.

Solution

1. We evaluate $T(u)$ using the defining formula:

$$T\begin{pmatrix} 1 \\ \pi \end{pmatrix} = \begin{pmatrix} \ln(1) \\ \cos(\pi) \\ \ln(1) \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}.$$

2. We have

$$T\begin{pmatrix} e^7 \\ 2\pi n \\ e^7 \end{pmatrix} = \begin{pmatrix} \ln(e^7) \\ \cos(2\pi n) \\ \ln(e^7) \end{pmatrix} = \begin{pmatrix} 7 \\ 1 \\ 7 \end{pmatrix}$$

   for any whole number n. Hence there are infinitely many such vectors.
3. The vector b from the previous part is an example of such a vector.
4. Since $\cos(y)$ is always between -1 and 1, the vector

$$w = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$$

   is not in the range of T.

This page titled 5.3: Matrix Transformations is shared under a GNU FDL license and was authored, remixed, and/or curated by Dan Margalit and Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- 3.1: Matrix Transformations by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

5.4: One-to-one and Onto Transformations

Objectives

1. Understand the definitions of one-to-one and onto transformations.
2. *Recipes:* verify whether a matrix transformation is one-to-one and/or onto.
3. *Pictures:* examples of matrix transformations that are/are not one-to-one and/or onto.
4. *Vocabulary words:* ***one-to-one***, ***onto***.

In this section, we discuss two of the most basic questions one can ask about a transformation: whether it is *one-to-one* and/or *onto*. For a matrix transformation, we translate these questions into the language of matrices.

One-to-one Transformations

Definition 5.4.1: One-to-one transformations

A transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is ***one-to-one*** if, for every vector b in $\mathbb{R}^m$, the equation $T(x) = b$ has *at most one* solution x in $\mathbb{R}^n$.

Remark

Another word for *one-to-one* is *injective*.

Here are some equivalent ways of saying that T is one-to-one:

- For every vector b in $\mathbb{R}^m$, the equation $T(x) = b$ has *zero or one* solution x in $\mathbb{R}^n$.
- Different inputs of T have different outputs.
- If $T(u) = T(v)$ then $u = v$.

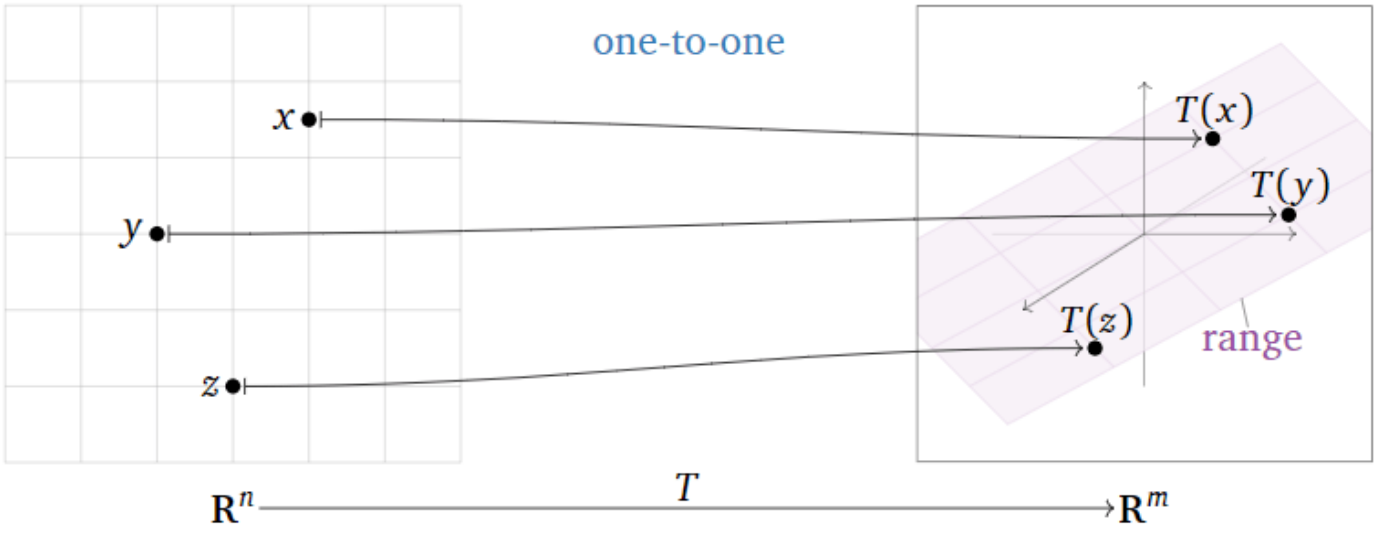


Figure 5.4.1

Here are some equivalent ways of saying that T is *not* one-to-one:

- There exists some vector b in $\mathbb{R}^m$ such that the equation $T(x) = b$ has *more than one* solution x in $\mathbb{R}^n$.
- There are two different inputs of T with the same output.
- There exist vectors u, v such that $u \neq v$ but $T(u) = T(v)$.

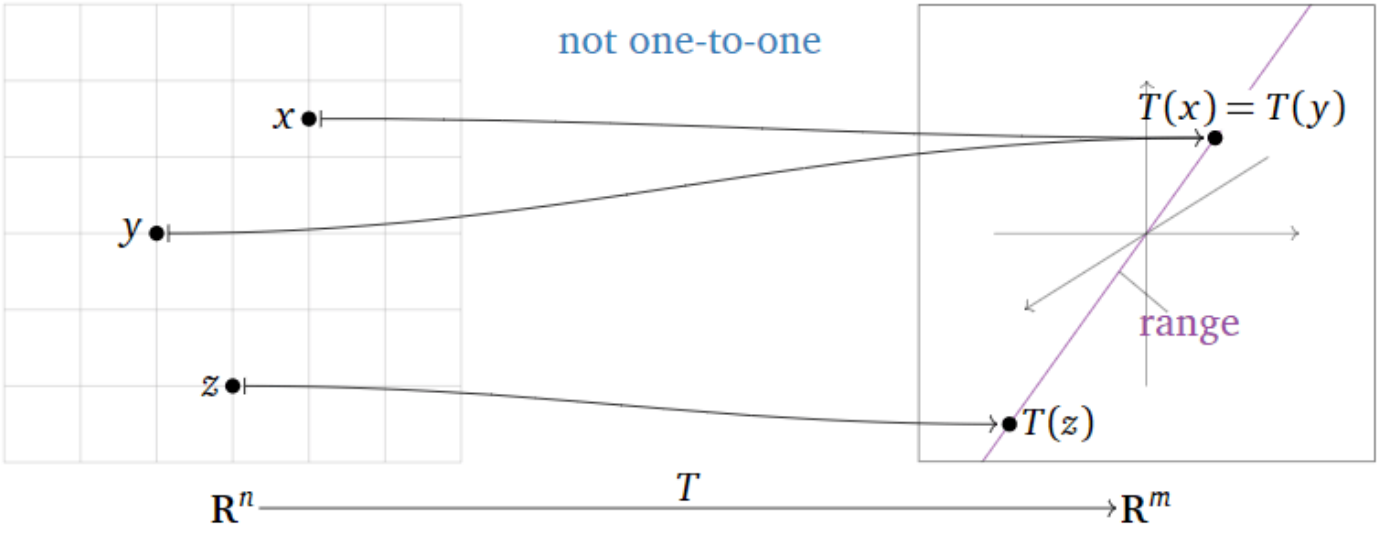


Figure 5.4.2

Example 5.4.1: Functions of one variable

The function $\sin\colon \mathbb{R} \to \mathbb{R}$ is not one-to-one. Indeed, $\sin(0) = \sin(\pi) = 0$, so the inputs 0 and π have the same output 0. In fact, the equation $\sin(x) = 0$ has infinitely many solutions $\ldots, -2\pi, -\pi, 0, \pi, 2\pi, \ldots$.

The function $\exp\colon \mathbb{R} \to \mathbb{R}$ defined by $\exp(x) = e^x$ is one-to-one. Indeed, if $T(x) = T(y)$, then $e^x = e^y$, so $\ln(e^x) = \ln(e^y)$, and hence $x = y$. The equation $T(x) = C$ has one solution $x = \ln(C)$ if $C > 0$, and it has zero solutions if $C \leq 0$.

The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is one-to-one. Indeed, if $f(x) = f(y)$ then $x^3 = y^3$; taking cube roots gives $x = y$. In other words, the only solution of $f(x) = C$ is $x = \sqrt[3]{C}$.

The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3 - x$ is not one-to-one. Indeed, $f(0) = f(1) = f(-1) = 0$, so the inputs $0, 1, -1$ all have the same output 0. The solutions of the equation $x^3 - x = 0$ are exactly the roots of $f(x) = x(x-1)(x+1)$, and this equation has three roots.

The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is not one-to-one. Indeed, $f(1) = 1 = f(-1)$, so the inputs 1 and -1 have the same outputs. The function $g\colon \mathbb{R} \to \mathbb{R}$ defined by $g(x) = |x|$ is not one-to-one for the same reason.

Example 5.4.2: A real-word transformation: robotics

Suppose you are building a robot arm with three joints that can move its hand around a plane, as in Example 3.1.11 in Section 3.1.

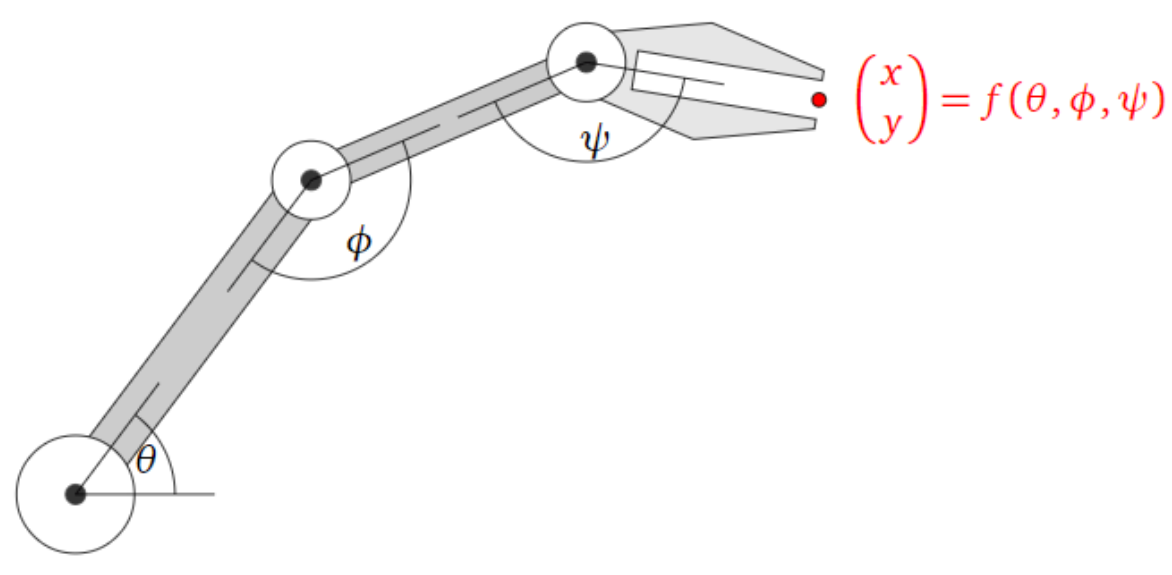


Figure 5.4.3

Define a transformation $f: \mathbb{R}^3 \to \mathbb{R}^2$ as follows: $f(\theta, \phi, \psi)$ is the (x, y) position of the hand when the joints are rotated by angles θ, ϕ, ψ, respectively. Asking whether f is one-to-one is the same as asking whether there is more than one way to move the arm in order to reach your coffee cup. (There is.)

Theorem 5.4.1: One-to-One Matrix Transformations

Let A be an $m \times n$ matrix, and let $T(x) = Ax$ be the associated matrix transformation. The following statements are equivalent:

1. T is one-to-one.
2. For every b in $\mathbb{R}^m$, the equation $T(x) = b$ has at most one solution.
3. For every b in $\mathbb{R}^m$, the equation $Ax = b$ has a unique solution or is inconsistent.
4. $Ax = 0$ has *only* the trivial solution.
5. The columns of A are linearly independent.
6. A has a pivot in every column.
7. The range of T has dimension n.

Proof available at the end of this section.

Recall that *equivalent* means that, for a given matrix, either all of the statements are true simultaneously, or they are all false.

Example 5.4.3: A matrix transformation that is one-to-one

Let A be the matrix

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

and define $T: \mathbb{R}^2 \to \mathbb{R}^3$ by $T(x) = Ax$. Is T one-to-one?

Solution

The reduced row echelon form of A is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Hence A has a pivot in every column, so T is one-to-one.

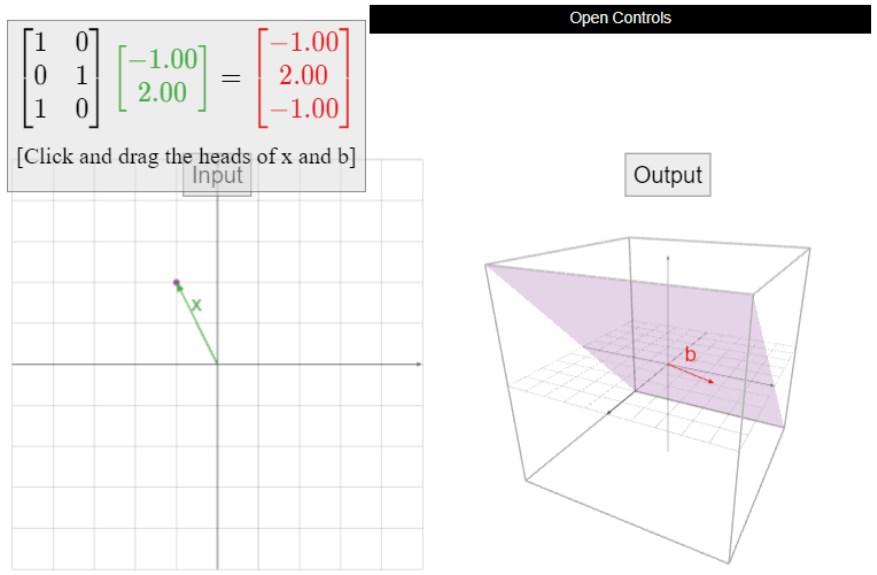


Figure 5.4.4: A picture of the matrix transformation T. As you drag the input vector on the left side, you see that different input vectors yield different output vectors on the right side.

Example 5.4.4: A matrix transformation that is not one-to-one

Let

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

and define $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ by $T(x) = Ax$. Is T one-to-one? If not, find two different vectors u, v such that $T(u) = T(v)$.

Solution

The matrix A is already in reduced row echelon form. It does not have a pivot in every column, so T is not one-to-one. Therefore, we know from the Theorem 5.4.1 that $Ax = 0$ has nontrivial solutions. If v is a nontrivial (i.e., nonzero) solution of $Av = 0$, then $T(v) = Av = 0 = A0 = T(0)$, so 0 and v are different vectors with the same output. For instance,

$$T\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 0 = T\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Geometrically, T is projection onto the xy-plane. Any two vectors that lie on the same vertical line will have the same projection. For b on the xy-plane, the solution set of $T(x) = b$ is the entire vertical line containing b. In particular, $T(x) = b$ has *infinitely many* solutions.

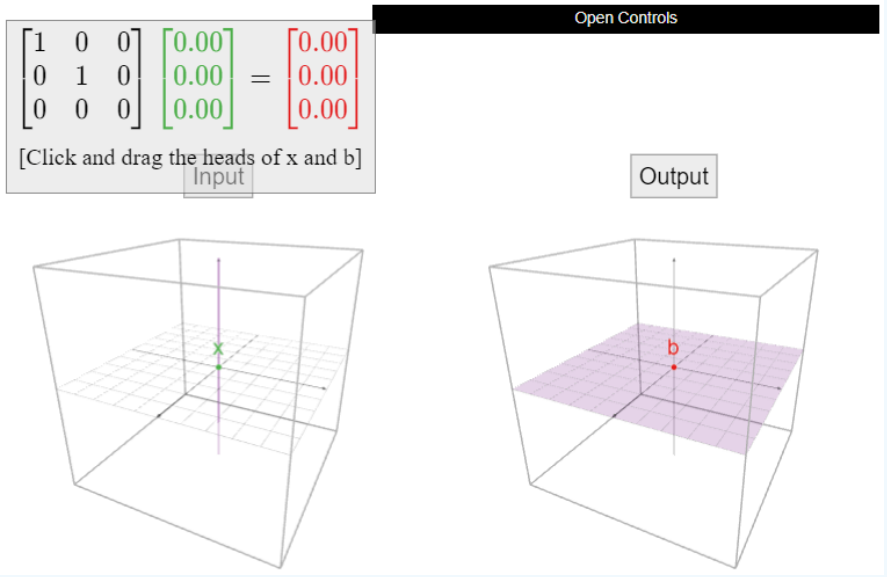


Figure 5.4.5: A picture of the matrix transformation T. The transformation T projects a vector onto the xy-plane. The violet line is the solution set of $T(x) = 0$. If you drag x along the violet line, the output $T(x) = Ax$ does not change. This demonstrates that $T(x) = 0$ has more than one solution, so T is not one-to-one.

Example 5.4.5: A matrix transformation that is not one-to-one

Let A be the matrix

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

and define $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ by $T(x) = Ax$. Is T one-to-one? If not, find two different vectors u, v such that $T(u) = T(v)$.

Solution

The reduced row echelon form of A is

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

There is not a pivot in every column, so T is not one-to-one. Therefore, we know from Theorem 5.4.1 that $Ax = 0$ has nontrivial solutions. If v is a nontrivial (i.e., nonzero) solution of $Av = 0$, then $T(v) = Av = 0 = A0 = T(0)$, so 0 and v are different vectors with the same output. In order to find a nontrivial solution, we find the parametric form of the solutions of $Ax = 0$ using the reduced matrix above:

$$\begin{cases} x \quad - \quad z = 0 \\ \quad y + z = 0 \end{cases} \implies \begin{cases} x = z \\ y = -z \end{cases}$$

The free variable is z. Taking $z = 1$ gives the nontrivial solution

$$T\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = 0 = T\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

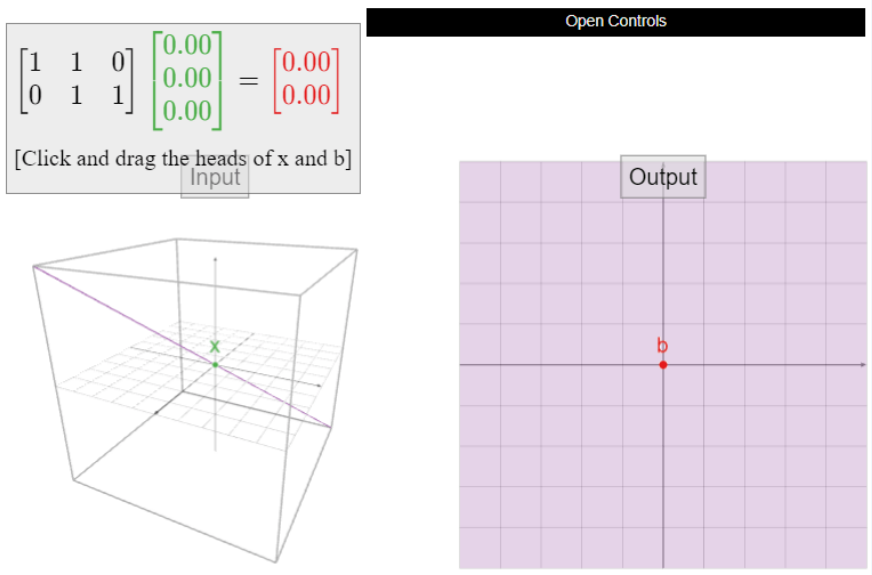


Figure 5.4.6: A picture of the matrix transformation T. The violet line is the null space of A, i.e., solution set of $T(x)=0$. If you drag x along the violet line, the output $T(x)=Ax$ does not change. This demonstrates that $T(x)=0$ has more than one solution, so T is not one-to-one.

Example 5.4.6: A matrix transformation that is not one-to-one

Let

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

and define $T\colon\mathbb{R}^3\to\mathbb{R}^2$ by $T(x)=Ax$. Is T one-to-one? If not, find two different vectors u,v such that $T(u)=T(v)$.

Solution

The reduced row echelon form of A is

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

There is not a pivot in every column, so T is not one-to-one. Therefore, we know from Theorem 5.4.1 that $Ax=0$ has nontrivial solutions. If v is a nontrivial (i.e., nonzero) solution of $Av=0$, then $T(v)=Av=0=A0=T(0)$, so 0 and v are different vectors with the same output. In order to find a nontrivial solution, we find the parametric form of the solutions of $Ax=0$ using the reduced matrix above:

$$x-y+2z=0\quad\Longrightarrow\quad x=y-2z.$$

The free variables are y and z. Taking $y=1$ and $z=0$ gives the nontrivial solution

$$T\begin{pmatrix}1\\1\\0\end{pmatrix}=\begin{pmatrix}1&-1&2\\-2&2&-4\end{pmatrix}\begin{pmatrix}1\\1\\0\end{pmatrix}=0=T\begin{pmatrix}0\\0\\0\end{pmatrix}.$$

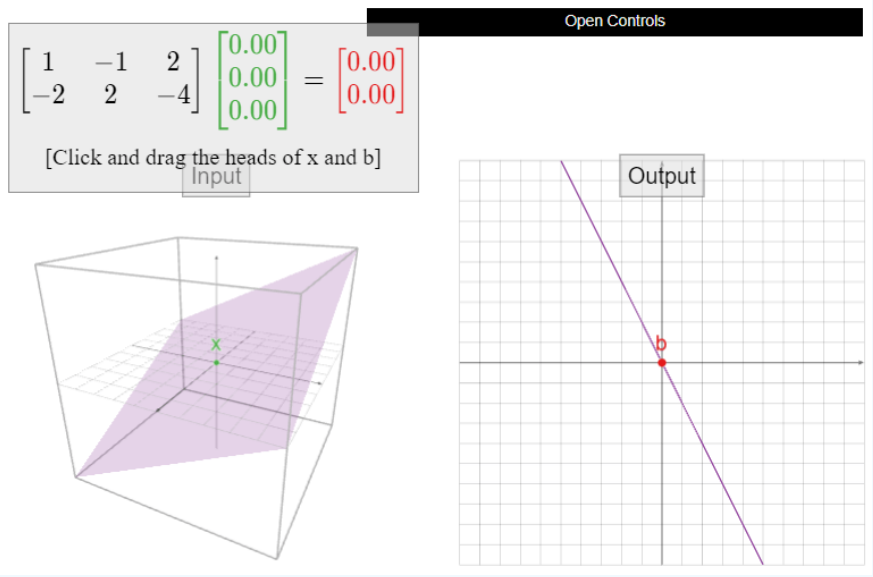


Figure 5.4.7: A picture of the matrix transformation T. The violet plane is the solution set of $T(x) = 0$. If you drag x along the violet plane, the output $T(x) = Ax$ does not change. This demonstrates that $T(x) = 0$ has more than one solution, so T is not one-to-one.

The previous three examples can be summarized as follows. Suppose that $T(x) = Ax$ is a matrix transformation that is *not* one-to-one. By Theorem 5.4.1, there is a nontrivial solution of $Ax = 0$. This means that the null space of A is not the zero space. All of the vectors in the null space are solutions to $T(x) = 0$. If you compute a nonzero vector v in the null space (by row reducing and finding the parametric form of the solution set of $Ax = 0$, for instance), then v and 0 both have the same output: $T(v) = Av = 0 = T(0)$.

Note 5.4.1: Wide matrices do not have one-to-one transformations

If $T: \mathbb{R}^n \to \mathbb{R}^m$ is a one-to-one matrix transformation, what can we say about the relative sizes of n and m?

The matrix associated to T has n columns and m rows. Each row and each column can only contain one pivot, so in order for A to have a pivot in every column, it must have *at least as many rows* as columns: $n \leq m$.

This says that, for instance, $\mathbb{R}^3$ is "too big" to admit a one-to-one linear transformation into $\mathbb{R}^2$.

Note that there exist tall matrices that are not one-to-one: for example,

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

does not have a pivot in every column.

Onto Transformations

Definition 5.4.2: Onto transformations

A transformation $T: \mathbb{R}^n \to \mathbb{R}^m$ is ***onto*** if, for every vector b in $\mathbb{R}^m$, the equation $T(x) = b$ has *at least one* solution x in $\mathbb{R}^n$.

Remark

Another word for *onto* is *surjective*.

Here are some equivalent ways of saying that T is onto:

- The range of T is equal to the codomain of T.
- Every vector in the codomain is the output of some input vector.

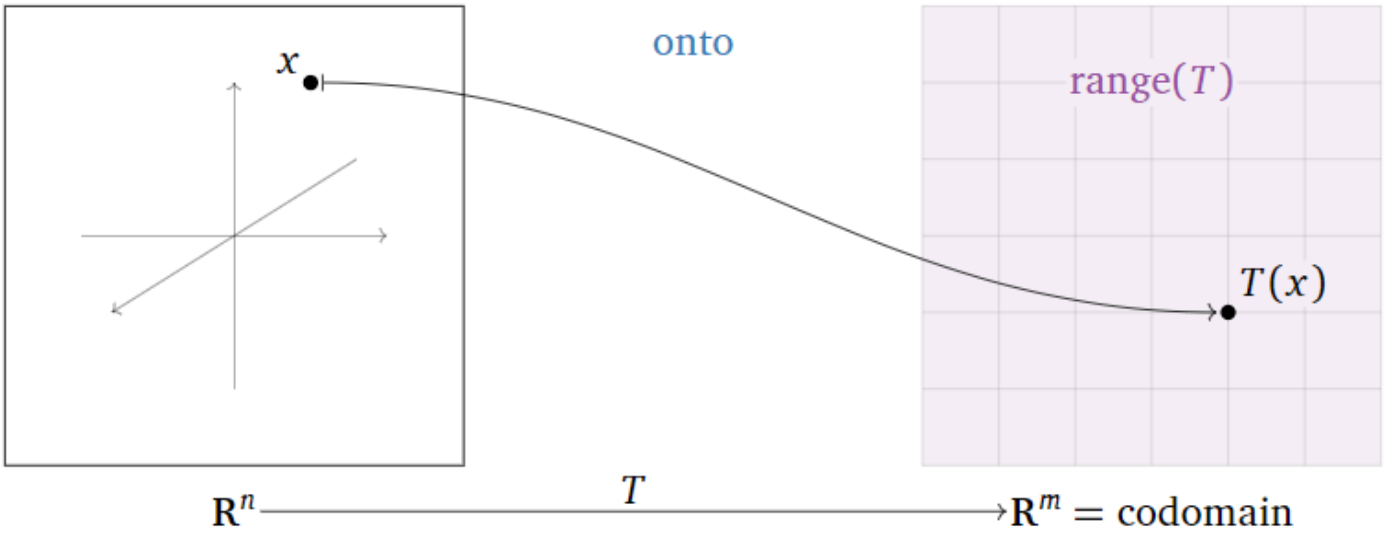


Figure 5.4.8

Here are some equivalent ways of saying that T is *not* onto:

- The range of T is smaller than the codomain of T.
- There exists a vector b in $\mathbb{R}^m$ such that the equation $T(x) = b$ does not have a solution.
- There is a vector in the codomain that is not the output of any input vector.

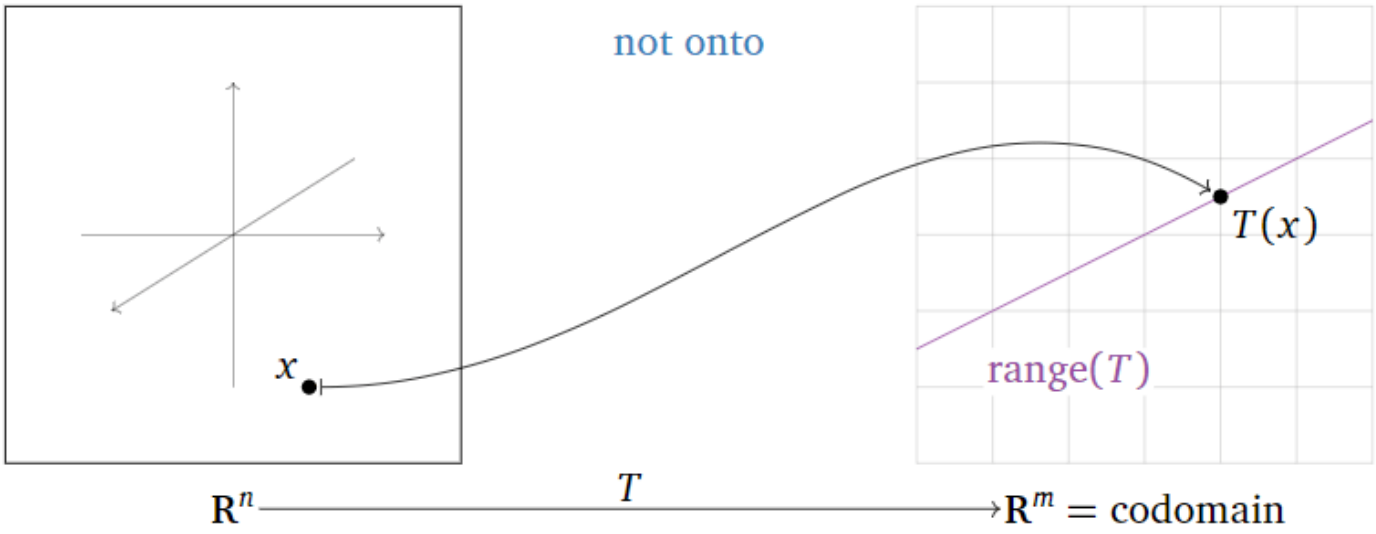


Figure 5.4.9

Example 5.4.7: Functions of one variable

The function $\sin\colon \mathbb{R} \to \mathbb{R}$ is not onto. Indeed, taking $b = 2$, the equation $\sin(x) = 2$ has no solution. The range of $\sin$ is the closed interval $[-1, 1]$, which is smaller than the codomain $\mathbb{R}$.

The function $\exp\colon \mathbb{R} \to \mathbb{R}$ defined by $\exp(x) = e^x$ is not onto. Indeed, taking $b = -1$, the equation $\exp(x) = e^x = -1$ has no solution. The range of $\exp$ is the set $(0, \infty)$ of all *positive* real numbers.

The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is onto. Indeed, the equation $f(x) = x^3 = b$ always has the solution $x = \sqrt[3]{b}$.

The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3 - x$ is onto. Indeed, the solutions of the equation $f(x) = x^3 - x = b$ are the roots of the polynomial $x^3 - x - b$; as this is a cubic polynomial, it has at least one real root.

Example 5.4.8: A real-word transformation: robotics

The robot arm transformation of Example 5.4.2 is not onto. The robot cannot reach objects that are very far away.

Theorem 5.4.2: Onto Matrix Transformations

Let A be an $m \times n$ matrix, and let $T(x) = Ax$ be the associated matrix transformation. The following statements are equivalent:

1. T is onto.
2. $T(x) = b$ has at least one solution for every b in $\mathbb{R}^m$.
3. $Ax = b$ is consistent for every b in $\mathbb{R}^m$.
4. The columns of A span $\mathbb{R}^m$.
5. A has a pivot in every row.
6. The range of T has dimension m.

Proof available at the end of this section.

Example 5.4.9: A matrix transformation that is onto

Let A be the matrix

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

and define $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ by $T(x) = Ax$. Is T onto?

Solution

The reduced row echelon form of A is

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

Hence A has a pivot in every row, so T is onto.

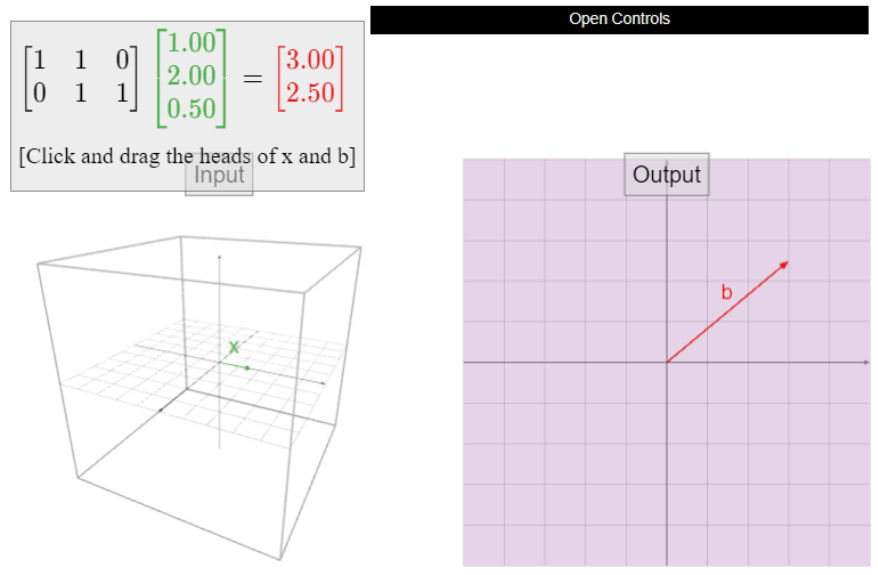


Figure 5.4.10: A picture of the matrix transformation T. Every vector on the right side is the output of T for a suitable input. If you drag b, the demo will find an input vector x with output b.

Example 5.4.10: A matrix transformation that is not onto

Let A be the matrix

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

and define $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ by $T(x) = Ax$. Is T onto? If not, find a vector b in $\mathbb{R}^3$ such that $T(x) = b$ has no solution.

Solution

The reduced row echelon form of A is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Hence A does not have a pivot in every row, so T is not onto. In fact, since

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \\ x \end{pmatrix},$$

we see that for every output vector of T, the third entry is equal to the first. Therefore,

$$b = (1, 2, 3)$$

is not in the range of T.

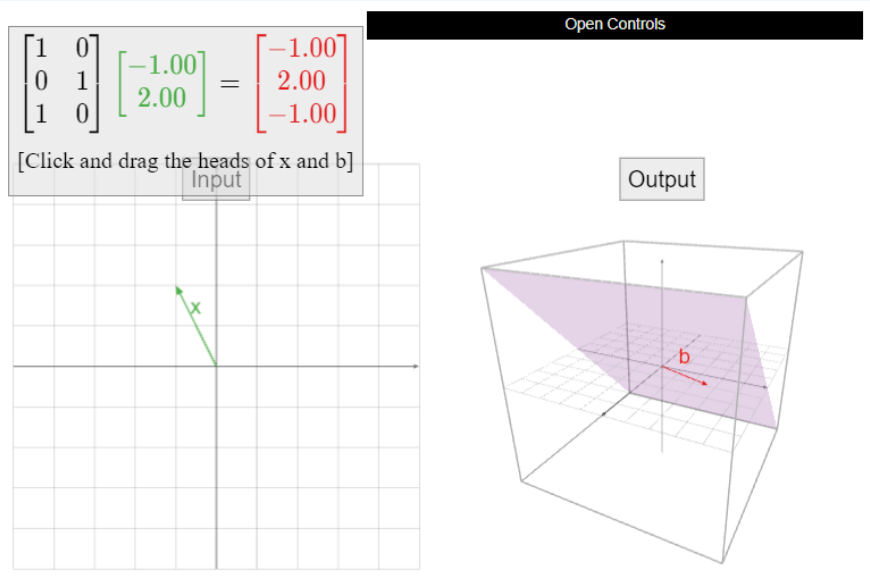


Figure 5.4.11: A picture of the matrix transformation T. The range of T is the violet plane on the right; this is smaller than the codomain $\mathbb{R}^3$. If you drag b off of the violet plane, then the equation $Ax = b$ becomes inconsistent; this means $T(x) = b$ has no solution.

Example 5.4.11: A matrix transformation that is not onto

Let

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

and define $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ by $T(x) = Ax$. Is T onto? If not, find a vector b in $\mathbb{R}^2$ such that $T(x) = b$ has no solution.

Solution

The reduced row echelon form of A is

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

There is not a pivot in every row, so T is not onto. The range of T is the column space of A, which is equal to

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

since all three columns of A are collinear. Therefore, any vector not on the line through $\begin{pmatrix}1\\-2\end{pmatrix}$ is not in the range of T. For instance, if $b=\begin{pmatrix}1\\1\end{pmatrix}$ then $T(x)=b$ has no solution.

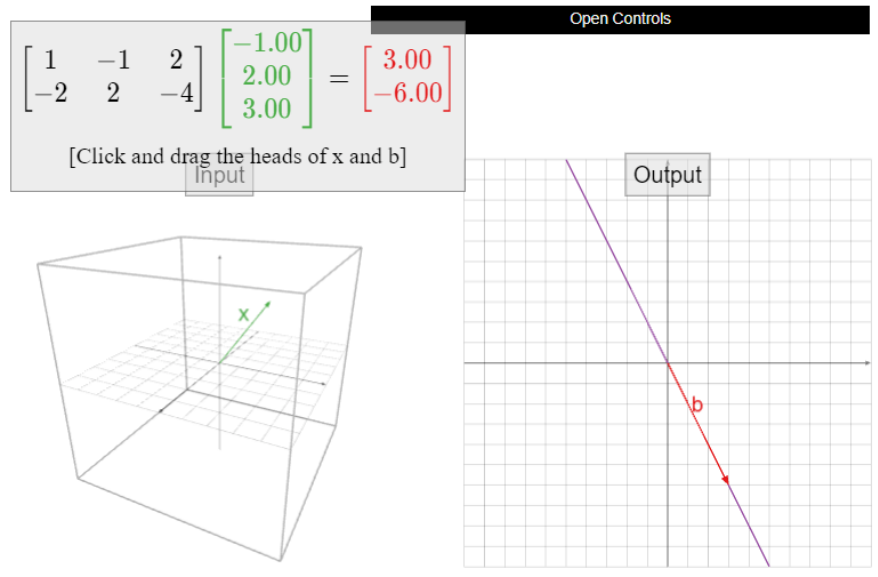


Figure 5.4.12: A picture of the matrix transformation T. The range of T is the violet line on the right; this is smaller than the codomain $\mathbb{R}^2$. If you drag b off of the violet line, then the equation $Ax=b$ becomes inconsistent; this means $T(x)=b$ has no solution.

The previous two examples illustrate the following observation. Suppose that $T(x)=Ax$ is a matrix transformation that is *not* onto. This means that $\text{range}(T)=\text{Col}(A)$ is a subspace of $\mathbb{R}^m$ of dimension less than m: perhaps it is a line in the plane, or a line in 3-space, or a plane in 3-space, etc. Whatever the case, the range of T is *very small* compared to the codomain. To find a vector not in the range of T, choose a random nonzero vector b in $\mathbb{R}^m$; you have to be extremely unlucky to choose a vector that is in the range of T. Of course, to check whether a given vector b is in the range of T, you have to solve the matrix equation $Ax=b$ to see whether it is consistent.

Note 5.4.2: Tall matrices do not have onto transformations

If $T\colon\mathbb{R}^n\to\mathbb{R}^m$ is an onto matrix transformation, what can we say about the relative sizes of n and m?

The matrix associated to T has n columns and m rows. Each row and each column can only contain one pivot, so in order for A to have a pivot in every row, it must have *at least as many columns* as rows: $m\leq n$.

This says that, for instance, $\mathbb{R}^2$ is "too small" to admit an onto linear transformation to $\mathbb{R}^3$.

Note that there exist wide matrices that are not onto: for example,

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

does not have a pivot in every row.

Comparison

The above expositions of one-to-one and onto transformations were written to mirror each other. However, "one-to-one" and "onto" are complementary notions: neither one implies the other. Below we have provided a chart for comparing the two. In the

chart, A is an $m \times n$ matrix, and $T: \mathbb{R}^n \to \mathbb{R}^m$ is the matrix transformation $T(x) = Ax$.

Table 5.4.1

T is one-to-one	T is onto
$T(x) = b$ has *at most one solution* for every b.	$T(x) = b$ has *at least one solution* for every b.
The columns of A are linearly independent.	The columns of A span $\mathbb{R}^m$.
A has a pivot column in every column.	A has a pivot in every row.
The range of T has dimension n.	The range of T has dimension m.

Example 5.4.12: Functions of one variable

The function $\sin: \mathbb{R} \to \mathbb{R}$ is neither one-to-one nor onto.

The function $\exp: \mathbb{R} \to \mathbb{R}$ defined by $\exp(x) = e^x$ is one-to-one but not onto.

The function $f: \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is one-to-one and onto.

The function $f: \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3 - x$ is onto but not one-to-one.

Example 5.4.13: A matrix transformation that is neither one-to-one nor onto

Let

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

and define $T: \mathbb{R}^3 \to \mathbb{R}^2$ by $T(x) = Ax$. This transformation is neither one-to-one nor onto, as we saw in Example 5.4.6 and Example 5.4.11.

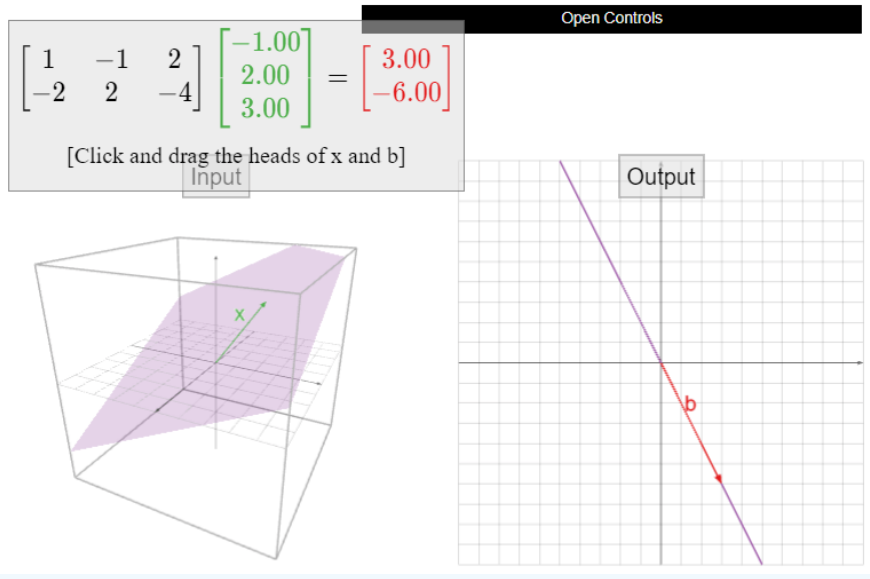


Figure 5.4.13: A picture of the matrix transformation T. The violet plane is the solution set of $T(x) = 0$. If you drag x along the violet plane, the output $T(x) = Ax$ does not change. This demonstrates that $T(x) = 0$ has more than one solution, so T is not one-to-one. The range of T is the violet line on the right; this is smaller than the codomain $\mathbb{R}^2$. If you drag b off of the violet line, then the equation $Ax = b$ becomes inconsistent; this means $T(x) = b$ has no solution.

Example 5.4.14: A matrix transformation that is one-to-one but not onto

Let A be the matrix

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

and define $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ by $T(x) = Ax$. This transformation is one-to-one but not onto, as we saw in this Example 5.4.3 and this Example 5.4.10.

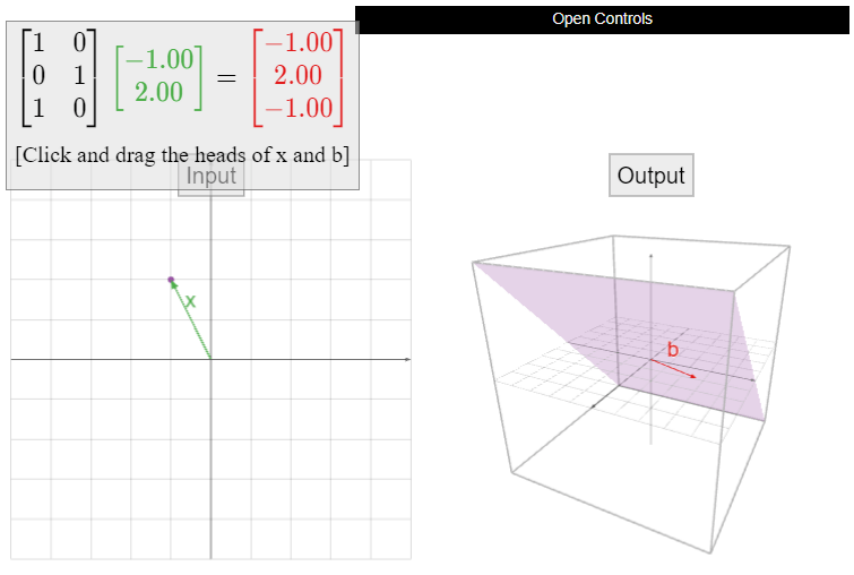


Figure 5.4.14: A picture of the matrix transformation T. The range of T is the violet plane on the right; this is smaller than the codomain $\mathbb{R}^3$. If you drag b off of the violet plane, then the equation $Ax = b$ becomes inconsistent; this means $T(x) = b$ has no solution. However, for b lying on the violet plane, there is a unique vector x such that $T(x) = b$.

Example 5.4.15: A matrix transformation that is onto but not one-to-one

Let A be the matrix

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

and define $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ by $T(x) = Ax$. This transformation is onto but not one-to-one, as we saw in Example 5.4.9.

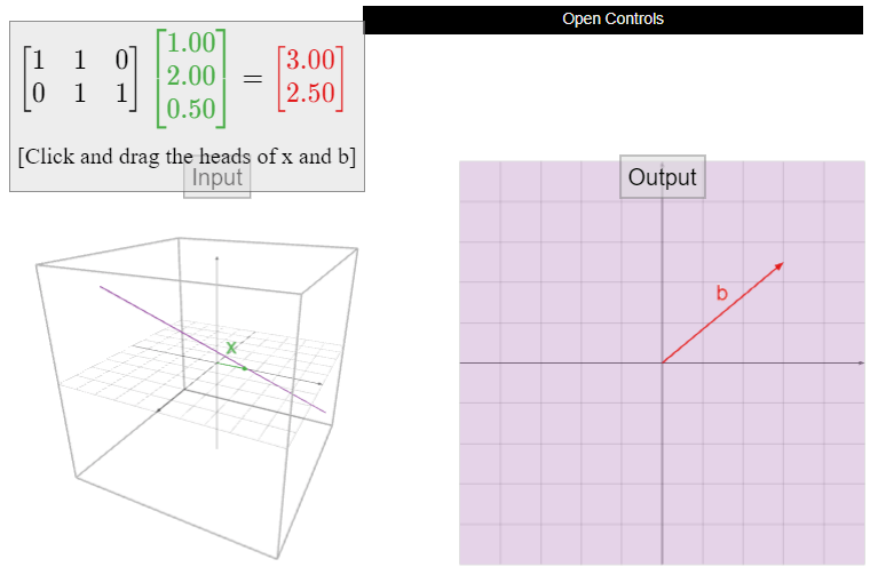


Figure 5.4.15: A picture of the matrix transformation T. Every vector on the right side is the output of T for a suitable input. If you drag b, the demo will find an input vector x with output b. The violet line is the null space of A, i.e., solution set of $T(x) = 0$. If you drag x along the violet line, the output $T(x) = Ax$ does not change. This demonstrates that $T(x) = 0$ has more than one solution, so T is not one-to-one.

Example 5.4.16: Matrix transformations that are both one-to-one and onto

In Subsection Matrices as Functions in Section 3.1, we discussed the transformations defined by several 2×2 matrices, namely:

$$\begin{aligned}
\text{Reflection:} \quad & A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \\
\text{Dilation:} \quad & A = \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix} \\
\text{Identity:} \quad & A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
\text{Rotation:} \quad & A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\
\text{Shear:} \quad & A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.
\end{aligned}$$

In each case, the associated matrix transformation $T(x) = Ax$ is both one-to-one and onto. A 2×2 matrix A has a pivot in every row if and only if it has a pivot in every column (if and only if it has two pivots), so in this case, the transformation T is one-to-one if and only if it is onto. One can see geometrically that they are onto (what is the input for a given output?), or that they are one-to-one using the fact that the columns of A are not collinear.

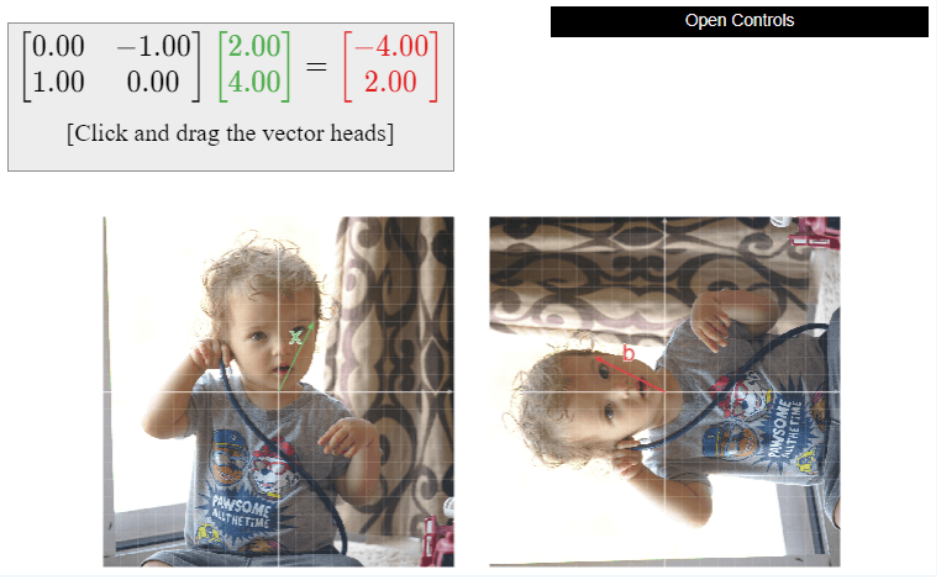


Figure 5.4.16: Counterclockwise rotation by 90° is a matrix transformation. This transformation is onto (if b is a vector in $\mathbb{R}^2$, then it is the output vector for the input vector which is b rotated *clockwise* by 90°), and it is one-to-one (different vectors rotate to different vectors).

Note 5.4.3: One-to-one is the same as onto for square matrices

We observed in the previous Example 5.4.16 that a square matrix has a pivot in every row if and only if it has a pivot in every column. Therefore, a matrix transformation T from $\mathbb{R}^n$ to itself is one-to-one if and only if it is onto: in this case, the two notions are equivalent.

Conversely, by Note 5.4.1 and Note 5.4.2, if a matrix transformation $T\colon \mathbb{R}^m \to \mathbb{R}^n$ is both one-to-one and onto, then $m = n$.

Note that in general, a transformation T is both one-to-one and onto if and only if $T(x) = b$ has *exactly one* solution for all b in $\mathbb{R}^m$.

Answers

Theorem 5.4.1: One-to-One Matrix Transformations

Statements 1, 2, and 3 are translations of each other. The equivalence of 3 and 4 follows from key observation 2.4.3 in Section 2.4: if $Ax = 0$ has only one solution, then $Ax = b$ has only one solution as well, or it is inconsistent. The equivalence of 4, 5, and 6 is a consequence of Recipe: Checking Linear Independence in Section 2.5, and the equivalence of 6 and 7 follows from the fact that the rank of a matrix is equal to the number of columns with pivots.

Theorem 5.4.2: Onto Matrix Transformations

Statements 1, 2, and 3 are translations of each other. The equivalence of 3, 4, 5, and 6 follows from Theorem 2.3.1 in Section 2.3.

This page titled 5.4: One-to-one and Onto Transformations is shared under a GNU FDL license and was authored, remixed, and/or curated by Dan Margalit and Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- 3.2: One-to-one and Onto Transformations by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

5.5: The Invertible Matrix Theorem

Objectives

1. *Theorem:* the invertible matrix theorem.

This section consists of a single important theorem containing many equivalent conditions for a matrix to be invertible. This is one of the most important theorems in this textbook. We will append two more criteria in Section 5.1.

Theorem 5.5.1: 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.

Proof available at the end of this section.

To reiterate, the invertible matrix theorem means:

Note 5.5.1

There are two kinds of *square* matrices:

1. invertible matrices, and
2. non-invertible matrices.

For invertible matrices, all of the statements of the invertible matrix theorem are true.

For non-invertible matrices, all of the statements of the invertible matrix theorem are false.

The reader should be comfortable translating any of the statements in the invertible matrix theorem into a statement about the pivots of a matrix.

Note 5.5.2: Other Conditions for Invertibility

The following conditions are also equivalent to the invertibility of a square matrix A. They are all simple restatements of conditions in the invertible matrix theorem.

1. The reduced row echelon form of A is the identity matrix I_n.
2. $Ax = 0$ has no solutions other than the trivial one.
3. $\text{nullity}(A) = 0$.
4. The columns of A form a basis for $\mathbb{R}^n$.

5. $Ax = b$ is consistent for all b in $\mathbb{R}^n$.
6. $\text{Col}(A) = \mathbb{R}^n$.
7. $\dim \text{Col}(A) = n$.
8. $\text{rank}(A) = n$.

Now we can show that to check $B = A^{-1}$, it's enough to show $AB = I_n$ *or* $BA = I_n$.

Corollary 5.5.1: A Left or Right Inverse Suffices

Let A be an $n \times n$ matrix, and suppose that there exists an $n \times n$ matrix B such that $AB = I_n$ *or* $BA = I_n$. Then A is invertible and $B = A^{-1}$.

Proof available at the end of this section.

We conclude with some common situations in which the invertible matrix theorem is useful.

Example 5.5.1

Is this matrix invertible?

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & 7 \\ -2 & -4 & 1 \end{pmatrix}$$

Solution

The second column is a multiple of the first. The columns are linearly dependent, so A does not satisfy condition 4 of the Theorem 5.5.1. Therefore, A is not invertible.

Example 5.5.2

Let A be an $n \times n$ matrix and let $T(x) = Ax$. Suppose that the range of T is $\mathbb{R}^n$. Show that the columns of A are linearly independent.

Solution

The range of T is the column space of A, so A satisfies condition 5 of the Theorem 5.5.1. Therefore, A also satisfies condition 4, which says that the columns of A are linearly independent.

Example 5.5.3

Let A be a 3×3 matrix such that

$$A \begin{pmatrix} 1 \\ 7 \\ 0 \end{pmatrix} = A \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}.$$

Show that the rank of A is at most 2.

Solution

If we set

$$b = A\begin{pmatrix} 1 \\ 7 \\ 0 \end{pmatrix} = A\begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix},$$

then $Ax = b$ has multiple solutions, so it does not satisfy condition 6 of the Theorem $5.5.1$. Therefore, it does not satisfy condition 5, so the columns of A do not span $\mathbb{R}^3$. Therefore, the column space has dimension strictly less than 3, the rank is at most 2.

Example $5.5.4$

Suppose that A is an $n \times n$ matrix such that $Ax = b$ is inconsistent some vector b. Show that $Ax = b$ has infinitely many solutions for some (other) vector b.

Solution

By hypothesis, A does not satisfy condition 6 of the Theorem $5.5.1$. Therefore, it does not satisfy condition 3, so $\mathrm{Nul}(A)$ is an infinite set. If we take $b = 0$, then the equation $Ax = b$ has infinitely many solutions.

Answers

Theorem $5.5.1$: Invertible Matrix Theorem

$(1 \leftrightarrow 2)$: The matrix A has n pivots if and only if its reduced row echelon form is the identity matrix I_n. This happens exactly when the procedure in Section 3.5, Theorem 3.5.1, to compute the inverse succeeds.

$(2 \leftrightarrow 3)$: The null space of a matrix is $\{0\}$ if and only if the matrix has no free variables, which means that every column is a pivot column, which means A has n pivots. See Recipe: Compute a Spanning Set for a Null Space in Section 2.6.

$(2 \leftrightarrow 4,\ 2 \leftrightarrow 5)$: These follow from the Recipe: Checking Linear Independence in Section 2.5 and Theorem 2.3.1, in Section 2.3, respectively, since A has n pivots if and only if has a pivot in every row/column.

$(4 + 5 \leftrightarrow 6)$: We know $Ax = b$ has at least one solution for every b if and only if the columns of A span \(\mathbb{R}^n \) by Theorem 3.2.2 in Section 3.2, and $Ax = b$ has at most one solution for every b if and only if the columns of A are linearly independent by Theorem 3.2.1 in Section 3.2. Hence $Ax = b$ has exactly one solution for every b if and only if its columns are linearly independent and span $\mathbb{R}^n$.

$(1 \leftrightarrow 7)$: This is the content of Theorem 3.5.3 in Section 3.5.

$(7 \implies 8 + 9)$: See Proposition 3.5.2 in Section 3.5.

$(8 \leftrightarrow 4,\ 9 \leftrightarrow 5)$: See Theorem 3.2.2 in Section 3.2 and Theorem 3.2.1 in Section 3.2.

Corollary $5.5.1$: A Left or Right Inverse Suffices

Suppose that $AB = I_n$. We claim that $T(x) = Ax$ is onto. Indeed, for any b in $\mathbb{R}^n$, we have

$$b = I_n b = (AB)b = A(Bb),$$

so $T(Bb) = b$, and hence b is in the range of T. Therefore, A is invertible by the Theorem $5.5.1$. Since A is invertible, we have

$$A^{-1} = A^{-1}I_n = A^{-1}(AB) = (A^{-1}A)B = I_n B = B,$$

so $B = A^{-1}$.

Now suppose that $BA = I_n$. We claim that $T(x) = Ax$ is one-to-one. Indeed, suppose that $T(x) = T(y)$. Then $Ax = Ay$, so $BAx = BAy$. But $BA = I_n$, so $I_n x = I_n y$, and hence $x = y$. Therefore, A is invertible by the Theorem $5.5.1$. One shows that $B = A^{-1}$ as above.

This page titled 5.5: The Invertible Matrix Theorem is shared under a GNU FDL license and was authored, remixed, and/or curated by Dan Margalit and Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- 3.6: The Invertible Matrix Theorem by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

6: EIGENVALUES AND EIGENVECTORS

6.1: An introduction to Eigenvalues and Eigenvectors
6.2: Finding eigenvalues and eigenvectors
6.3: Diagonalization, similarity, and powers of a matrix
6.4: Dynamical Systems
6.5: Markov chains and Google's PageRank algorithm

This page titled 6: Eigenvalues and Eigenvectors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by David Austin via source content that was edited to the style and standards of the LibreTexts platform.

- 4: Eigenvalues and Eigenvectors by David Austin is licensed CC BY 4.0. Original source: https://davidaustinm.github.io/ula/ula.html.

6.1: An introduction to Eigenvalues and Eigenvectors

This section introduces the concept of eigenvalues and eigenvectors and offers an example that motivates our interest in them. The point here is to develop an intuitive understanding of eigenvalues and eigenvectors and explain how they can be used to simplify some problems that we have previously encountered. In the rest of this chapter, we will develop this concept into a richer theory and illustrate its use with more meaningful examples.

Preview Activity 4.1.1.

Before we introduce the definition of eigenvectors and eigenvalues, it will be helpful to remember some ideas we have seen previously.

1. Suppose that $\mathbf{v}$ is the vector shown in the figure. Sketch the vector $2\mathbf{v}$ and the vector $-\mathbf{v}$.

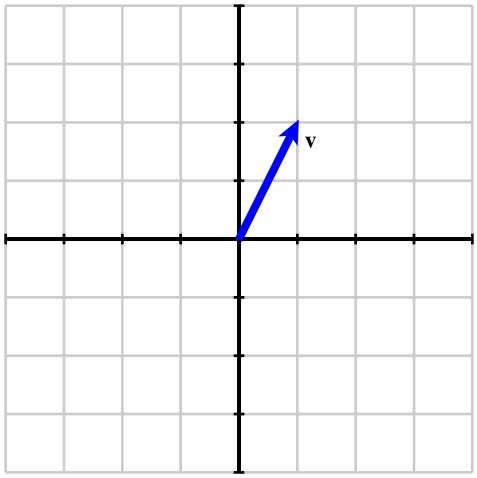


Figure 6.1.1

- State the geometric effect that scalar multiplication has on the vector $\mathbf{v}$. Then sketch all the vectors of the form $\lambda\mathbf{v}$ where λ is a scalar.
- State the geometric effect of the matrix transformation defined by

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

- Suppose that A is a 2×2 matrix and that $\mathbf{v}_1$ and $\mathbf{v}_2$ are vectors such that

$$A\mathbf{v}_1 = 3\mathbf{v}_1, \qquad A\mathbf{v}_2 = -\mathbf{v}_2.$$

  Use the linearity of matrix multiplication to express the following vectors in terms of $\mathbf{v}_1$ and $\mathbf{v}_2$.

  1. $A(4\mathbf{v}_1)$.
  2. $A(\mathbf{v}_1 + \mathbf{v}_2)$.
  3. $A(4\mathbf{v}_1 - 3\mathbf{v}_2)$.
  4. $A^2\mathbf{v}_1$.
  5. $A^2(4\mathbf{v}_1 - 3\mathbf{v}_2)$.
  6. $A^4\mathbf{v}_1$.

A few examples

We will now introduce the definition of eigenvalues and eigenvectors and then look at a few simple examples.

Definition 4.1.1

Given a square $n \times n$ matrix A, we say that a nonzero vector $\mathbf{v}$ is an *eigenvector* of A if there is a scalar λ such that

$$A\mathbf{v} = \lambda\mathbf{v}.$$

The scalar λ is called the *eigenvalue* associated to the eigenvector $\mathbf{v}$.

At first glance, there is a lot going on in this definition so let's look at an example.

Example 4.1.2

Consider the matrix $A = \begin{bmatrix} 7 & 6 \\ 6 & -2 \end{bmatrix}$ and the vector $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. We find that

$$A\mathbf{v} = \begin{bmatrix} 7 & 6 \\ 6 & -2 \end{bmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 20 \\ 10 \end{pmatrix} = 10 \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 10\mathbf{v}.$$

In other words, $A\mathbf{v} = 10\mathbf{v}$, which says that $\mathbf{v}$ is an eigenvector of the matrix A with associated eigenvalue $\lambda = 10$.

Similarly, if $\mathbf{w} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$, we find that

$$A\mathbf{w} = \begin{bmatrix} 7 & 6 \\ 6 & -2 \end{bmatrix} \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ -10 \end{pmatrix} = -5 \begin{pmatrix} -1 \\ 2 \end{pmatrix} = -5\mathbf{w}.$$

Here again, we have $A\mathbf{w} = -5\mathbf{w}$ showing that $\mathbf{w}$ is an eigenvector of A with associated eigenvalue $\lambda = -5$.

Activity 4.1.2.

This definition has an important geometric interpretation that we will investigate here.

1. Suppose that $\mathbf{v}$ is a nonzero vector and that λ is a scalar. What is the geometric relationship between $\mathbf{v}$ and $\lambda\mathbf{v}$?
2. Let's now consider the eigenvector condition: $A\mathbf{v} = \lambda\mathbf{v}$. Here we have two vectors, $\mathbf{v}$ and $A\mathbf{v}$. If $A\mathbf{v} = \lambda\mathbf{v}$, what is the geometric relationship between $\mathbf{v}$ and $A\mathbf{v}$?
3. Choose the matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. Move the vector $\mathbf{v}$ so that the eigenvector condition holds. What is the eigenvector $\mathbf{v}$ and what is the associated eigenvalue?
4. By algebraically computing $A\mathbf{v}$, verify that the eigenvector condition holds for the vector $\mathbf{v}$ that you found.
5. If you multiply the eigenvector $\mathbf{v}$ that you found by 2, do you still have an eigenvector? If so, what is the associated eigenvalue?
6. Are you able to find another eigenvector $\mathbf{v}$ that is not a scalar multiple of the first one that you found? If so, what is the eigenvector and what is the associated eigenvalue?
7. Now consider the matrix $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$. Use the diagram to describe any eigenvectors and associated eigenvalues.
8. Finally, consider the matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Use the diagram to describe any eigenvectors and associated eigenvalues. What geometric transformation does this matrix perform on vectors? How does this explain the presence of any eigenvectors?

Let's consider the ideas we saw in the activity in some more depth. To be an eigenvector of A, the vector $\mathbf{v}$ must satisfy $A\mathbf{v} = \lambda\mathbf{v}$

for some scalar λ. This means that $\mathbf{v}$ and $A\mathbf{v}$ are scalar multiples of one another, which means they must lie on the same line.

Consider now the matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. On the left of Figure 4.1.3, we see that $\mathbf{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is not an eigenvector of A since the vectors $\mathbf{v}$ and $A\mathbf{v}$ do not lie on the same line. On the right, however, we see that $\mathbf{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector. In fact, $A\mathbf{v}$ is obtained from $\mathbf{v}$ by stretching $\mathbf{v}$ by a factor of 3. Therefore, $\mathbf{v}$ is an eigenvector of A with eigenvalue $\lambda = 3$.

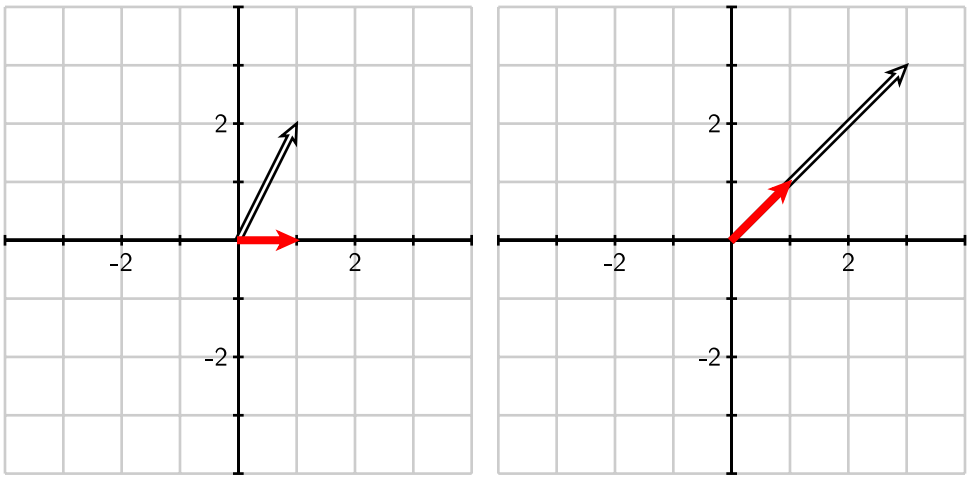


Figure Misplaced & : On the left, the vector $\mathbf{v}$ is not an eigenvector. On the right, the vector $\mathbf{v}$ is an eigenvector with eigenvalue $\lambda = 3$.

It is not difficult to see that any multiple of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is also an eigenvector of A with eigenvalue $\lambda = 3$. Indeed, we will see later that all the eigenvectors associated to a given eigenvalue form a subspace of $\mathbb{R}^n$.

In Figure 4.1.4, we see that $\mathbf{v} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ is also an eigenvector with eigenvalue $\lambda = -1$.

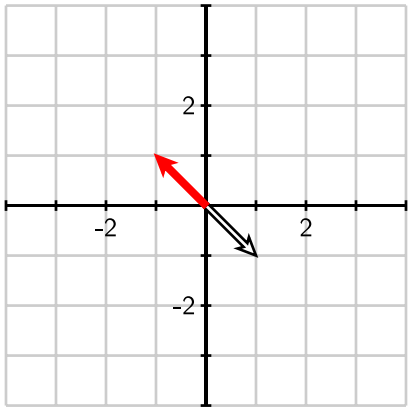


Figure 6.1.4: Here we see another eigenvector $\mathbf{v}$ with eigenvalue $\lambda = -1$.

The interactive diagram we used in the activity is meant to convey the fact that the eigenvectors of a matrix A are special vectors. Most of the time, the vectors $\mathbf{v}$ and $A\mathbf{v}$ appear somewhat unrelated. For certain vectors, however, $\mathbf{v}$ and $A\mathbf{v}$ line up with one another. Something important is going on when that happens so we call attention to those vectors by calling them eigenvectors. For these vectors, the operation of multiplying by A reduces to the much simpler operation of scalar multiplying by λ. The reason eigenvectors are important is because it is extremely convenient to be able to replace matrix multiplication by scalar multiplication.

Eigen is a German word that can be interpreted as meaning "characteristic". As we will see, the eigenvectors and eigenvalues of a matrix A give an important characterization of the matrix.

The usefulness of eigenvalues and eigenvectors

In the next section, we will introduce an algebraic technique for finding the eigenvalues and eigenvectors of a matrix. Before doing that, however, we would like to discuss why eigenvalues and eigenvectors are so useful.

Let's continue looking at the example $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. We have seen that $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector with eigenvalue $\lambda = 3$ and $\mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ is an eigenvector with eigenvalue $\lambda = -1$. This means that $A\mathbf{v}_1 = 3\mathbf{v}_1$ and $A\mathbf{v}_2 = -\mathbf{v}_2$. By the linearity of matrix multiplication, we can easily understand what happens when we multiply a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ by A:

$$A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = 3c_1\mathbf{v}_1 - c_2\mathbf{v}_2.$$

For instance, if we consider the vector $\mathbf{x} = \mathbf{v}_1 - 2\mathbf{v}_2$,
we find that

$$A\mathbf{x} = A(\mathbf{v}_1 - 2\mathbf{v}_2)$$
$$A\mathbf{x} = 3\mathbf{v}_1 + 2\mathbf{v}_2$$

as seen in the figure.

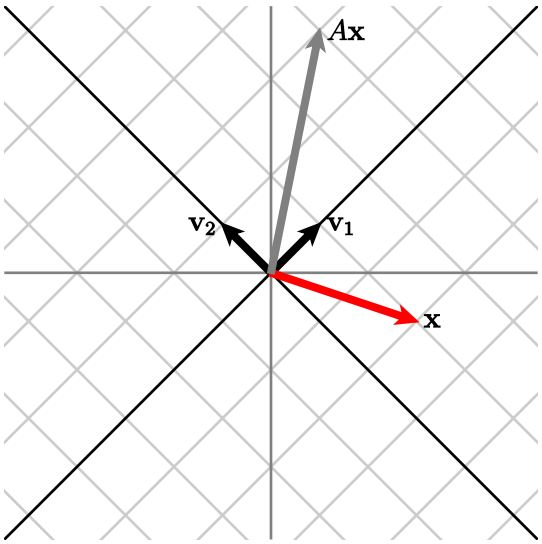


Figure 6.1.5

In other words, multiplying by A has the effect of stretching a vector $\mathbf{x}$ in the $\mathbf{v}_1$ direction by a factor of 3 and flipping the $\mathbf{x}$ in $\mathbf{v}_2$ direction.

We can draw an analogy with the more familiar example of the diagonal matrix $D = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}$. As we have seen, the effect of multiplying a vector $\mathbf{x}$ by D is to stretch $\mathbf{x}$ horizontally by a factor of 3 and to flip $\mathbf{x}$ vertically. This is illustrated in Figure 4.1.6.

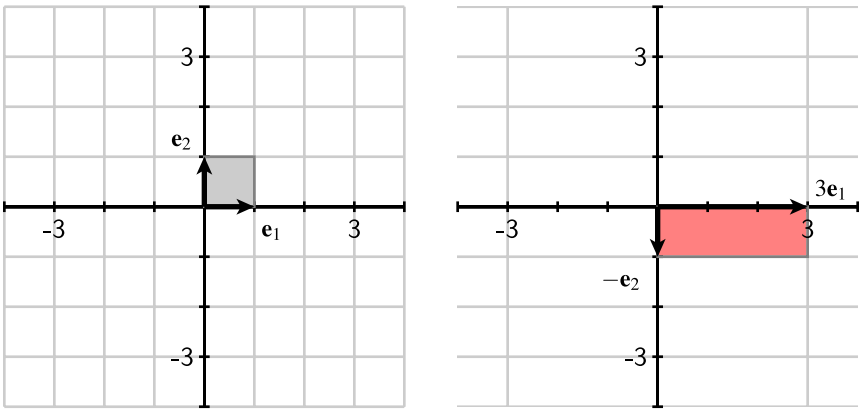


Figure 6.1.6: The diagonal matrix D stretches vectors horizontally by a factor of 3 and flips vectors vertically.

The matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ has a similar effect when viewed in the basis defined by the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$, as seen in Figure 4.1.7.

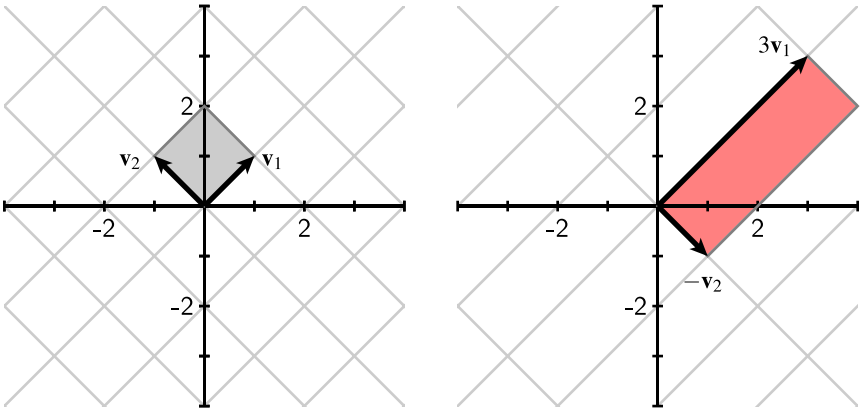


Figure 6.1.7: The matrix A has the same geometric effect as the diagonal matrix D when expressed in the coordinate system defined by the basis of eigenvectors.

In a sense that will be made precise later, having a set of eigenvectors of A that forms a basis of $\mathbb{R}^2$ enables us to think of A as being equivalent to a diagonal matrix D. Of course, as the other examples in the previous activity show, it may not always be possible to form a basis from the eigenvectors of a matrix. For example, the only eigenvectors of the matrix $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$, which represents a shear, have the form $\begin{pmatrix} x \\ 0 \end{pmatrix}$. In this example, we are not able to create a basis for $\mathbb{R}^2$ consisting of eigenvectors of the matrix. This is also true for the matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, which represents a 90° rotation.

Activity 4.1.3.

Let's consider an example that illustrates how we can put these ideas to use.

Suppose that we work for a car rental company that has two locations, P and Q. When a customer rents a car at one location, they have the option to return it to either location at the end of the day. After doing some market research, we determine:

- 80% of the cars rented at location P are returned to P and 20% are returned to Q.
- 40% of the cars rented at location Q are returned to Q and 60% are returned to P.

1. Suppose that there are 1000 cars at location P and no cars at location Q on Monday morning. How many cars are there are

locations P and Q at the end of the day on Monday?

2. How many are at locations P and Q at end of the day on Tuesday?
3. If we let P_k and Q_k be the number of cars at locations P and Q, respectively, at the end of day k, we then have

$$P_{k+1} = 0.8P_k + 0.6Q_k$$
$$Q_{k+1} = 0.2P_k + 0.4Q_k.$$

We can write the vector $\mathbf{x}_k = \begin{pmatrix} P_k \\ Q_k \end{pmatrix}$ to reflect the number of cars at the two locations at the end of day k, which says that

$$\mathbf{x}_{k+1} = \begin{bmatrix} 0.8 & 0.6 \\ 0.2 & 0.4 \end{bmatrix} \mathbf{x}_k$$

$$\text{or} \qquad \mathbf{x}_{k+1} = A\mathbf{x}_k$$

where $A = \begin{bmatrix} 0.8 & 0.6 \\ 0.2 & 0.4 \end{bmatrix}$.

Suppose that

$$\mathbf{v}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Compute $A\mathbf{v}_1$ and $A\mathbf{v}_2$ to demonstrate that $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of A. What are the associated eigenvalues λ_1 and λ_2?

4. We said that 1000 cars are initially at location P and none at location Q. This means that the initial vector describing the number of cars is $\mathbf{x}_0 = \begin{pmatrix} 1000 \\ 0 \end{pmatrix}$. Write $\mathbf{x}_0$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.
5. Remember that $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of A. Use the linearity of matrix multiplicaiton to write the vector $\mathbf{x}_1 = A\mathbf{x}_0$, describing the number of cars at the two locations at the end of the first day, as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.
6. Write the vector $\mathbf{x}_2 = A\mathbf{x}_1$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. Then write the next few vectors as linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$:
    1. $\mathbf{x}_3 = A\mathbf{x}_2$.
    2. $\mathbf{x}_4 = A\mathbf{x}_3$.
    3. $\mathbf{x}_5 = A\mathbf{x}_4$.
    4. $\mathbf{x}_6 = A\mathbf{x}_5$.
7. What will happen to the number of cars at the two locations after a very long time? Explain how writing $\mathbf{x}_0$ as a linear combination of eigenvectors helps you determine the long-term behavior.

This activity is important and motivates much of our work with eigenvalues and eigenvectors so we will review it now making sure we have a clear understanding of the concepts that arise.

First, we compute

$$A\mathbf{v}_1 = \begin{bmatrix} 0.8 & 0.6 \\ 0.2 & 0.4 \end{bmatrix} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix} = 1\mathbf{v}_1$$

$$A\mathbf{v}_2 = \begin{bmatrix} 0.8 & 0.6 \\ 0.2 & 0.4 \end{bmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.2 \\ 0.2 \end{pmatrix} = 0.2\mathbf{v}_2.$$

This shows that $\mathbf{v}_1$ is an eigenvector of A with eigenvalue $\lambda_1 = 1$ and $\mathbf{v}_2$ is an eigenvector of A with eigenvalue $\lambda_2 = 0.2$.

By the linearity of matrix matrix multiplication, we have

$$A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1\mathbf{v}_1 + 0.2c_2\mathbf{v}_2.$$

Therefore, we will write the vector describing the initial distribution of cars $\mathbf{x}_0 = \begin{pmatrix} 1000 \\ 0 \end{pmatrix}$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$; that is, $\mathbf{x}_0 = c_1\mathbf{v}_2 + c_2\mathbf{v}_2$. To do, we form the augmented matrix and row reduce:

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{x}_0 \end{bmatrix} = \begin{bmatrix} 3 & -1 & 1000 \\ 1 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 250 \\ 0 & 1 & -250 \end{bmatrix}.$$

Therefore, $\mathbf{x}_0 = 250\mathbf{v}_1 - 250\mathbf{v}_2$.

To determine the distribution of cars on subsequent days, we will repeatedly multiply by A. We find that

$$\begin{aligned}
\mathbf{x}_1 &= A\mathbf{x}_0 = A(250\mathbf{v}_1 - 250\mathbf{v}_2) = 250\mathbf{v}_1 - (0.2)250\mathbf{v}_2 \\
\mathbf{x}_2 &= A\mathbf{x}_1 = A(250\mathbf{v}_1 - (0.2)250\mathbf{v}_2) = 250\mathbf{v}_1 - (0.2)^2250\mathbf{v}_2 \\
\mathbf{x}_3 &= A\mathbf{x}_2 = A(250\mathbf{v}_1 - (0.2)^2250\mathbf{v}_2) = 250\mathbf{v}_1 - (0.2)^3250\mathbf{v}_2. \\
\mathbf{x}_4 &= A\mathbf{x}_3 = A(250\mathbf{v}_1 - (0.2)^3250\mathbf{v}_2) = 250\mathbf{v}_1 - (0.2)^4250\mathbf{v}_2 \\
\mathbf{x}_5 &= A\mathbf{x}_4 = A(250\mathbf{v}_1 - (0.2)^4250\mathbf{v}_2) = 250\mathbf{v}_1 - (0.2)^5250\mathbf{v}_2
\end{aligned}$$

In particular, this shows us that

$$\mathbf{x}_5 = 250\mathbf{v}_1 - (0.2)^5250\mathbf{v}_2 = \begin{pmatrix} 250 \cdot 3 - (0.2)^5250 \cdot (-1) \\ 250 \cdot 1 - (0.2)^5250 \cdot 1 \end{pmatrix} = \begin{pmatrix} 750.09 \\ 249.92 \end{pmatrix}.$$

Taking notice of the pattern, we may write

$$\mathbf{x}_k = 250\mathbf{v}_1 - (0.2)^k250\mathbf{v}_2.$$

Multiplying a number by 0.2 is the same as taking 20% of that number. As each day goes by, the second term is multiplied by 0.2 so the coefficient of $\mathbf{v}_2$ in the expression for $\mathbf{x}_k$ will eventually become extremely small. We therefore see that the distribution of cars will stabilize at $\mathbf{x} = 250\mathbf{v}_1 = \begin{pmatrix} 750 \\ 250 \end{pmatrix}$.

Notice how our use of the eigenvalues and eigenvectors of A enable us to look far into the future without having to repeatedly multiply a vector by the matrix A. Knowing the eigenvectors allows us to replace matrix multiplication with the simpler operation of scalar multiplication. This is a powerful tool that we will develop more in the rest of this chapter.

Notice also how this example relies on the fact that we can express the initial vector $\mathbf{x}_0$ as a linear combination of eigenvectors. For this reason, we would like, given an $n \times n$ matrix, to be able to create a basis of $\mathbb{R}^n$ of eigenvectors. We will frequently return to this question in later sections.

Question 4.1.7.

Given an $n \times n$ matrix A, when can we form a basis for $\mathbb{R}^n$ consisting of eigenvectors of A?

Summary

We defined an eigenvector of a square matrix A to be a nonzero vector $\mathbf{v}$ such that $A\mathbf{v} = \lambda\mathbf{v}$ for some scalar λ, which is called the eigenvalue associated to $\mathbf{v}$.

- If $\mathbf{v}$ is an eigenvector, then matrix multiplication by A reduces to the simpler operation of scalar multiplication by λ.
- Scalar multiples of an eigenvector are also eigenvectors. In fact, we will see that the eigenvectors associated to an eigenvalue λ form a subspace.
- If we can form a basis for $\mathbb{R}^n$ consisting of eigenvectors of A, then A is, in some sense, equivalent to a diagonal matrix.
- Rewriting a vector $\mathbf{x}$ as a linear combination of eigenvectors of A makes it easy to repeatedly multiply $\mathbf{x}$ by A.

Exercises 4.1.4Exercises

1

Consider the matrix and vectors

$$A = \begin{bmatrix} 8 & -10 \\ 5 & -7 \end{bmatrix}, \qquad \mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

1. Show that the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of A and find their associated eigenvalues.
2. Express the vector $\mathbf{x} = \begin{pmatrix} -4 \\ -1 \end{pmatrix}$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.
3. Use this expression to compute $A\mathbf{x}$, $A^2\mathbf{x}$, and $A^{-1}\mathbf{x}$ as a linear combination of eigenvectors.

2

Consider the matrix and vectors

$$A = \begin{bmatrix} -5 & -2 & 2 \\ 24 & 14 & -10 \\ 21 & 14 & -10 \end{bmatrix}, \qquad \mathbf{v}_1 = \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 2 \\ -3 \\ 0 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}$$

1. Show that the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are eigenvectors of A and find their associated eigenvalues.
2. Express the vector $\mathbf{x} = \begin{pmatrix} 0 \\ -3 \\ -4 \end{pmatrix}$ as a linear combination of the eigenvectors.
3. Use this expression to compute $A\mathbf{x}$, $A^2\mathbf{x}$, and $A^{-1}\mathbf{x}$ as a linear combination of eigenvectors.

3

Suppose that A is an $n \times n$ matrix.

1. Explain why $\lambda = 0$ is an eigenvalue if and only if there is a non-trivial solution to the homogeneous equation $A\mathbf{x} = 0$.
2. Explain why such a matrix A is not invertible if and only if $\lambda = 0$ is an eigenvalue.
3. If A is an invertible $n \times n$ matrix having an eigenvector $\mathbf{v}$ and associated eigenvalue λ, explain why $\mathbf{v}$ is also an eigenvector of A^{-1} with associated eigenvalue λ^{-1}.
4. If A is an $n \times n$ matrix with eigenvector $\mathbf{v}$ and associated eigenvalue λ, explain why $\mathbf{v}$ is also an eigenvector of A^2 with associated eigenvalue λ^2.
5. The matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ has eigenvectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ and associated eigenvalues $\lambda_1 = 3$ and $\lambda = -1$. What are some eigenvectors and associated eigenvalues for A^5?

4

Suppose that A is a matrix with eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ and eigenvalues $\lambda_1 = -1$ and $\lambda_2 = 2$ as shown.

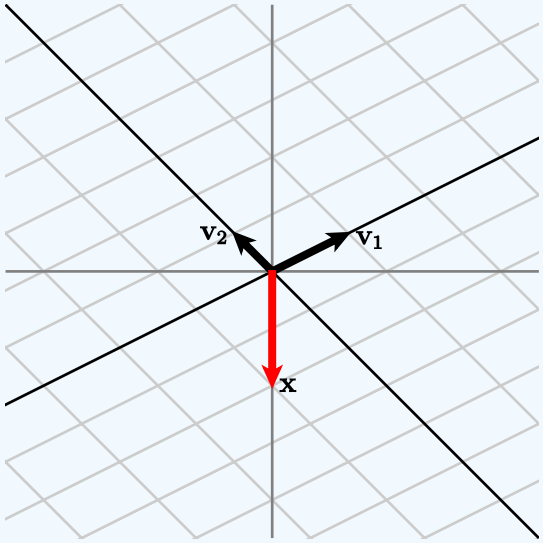


Figure 6.1.8: \text { The vectors } \mathbf{v}_1 \text { and } \mathbf{v}_2 \text { are eigenvectors of } A \text {. }

Sketch the vectors $A\mathbf{x}$, $A^2\mathbf{x}$, and $A^{-1}\mathbf{x}$.

5

For the following matrices, find the eigenvectors and associated eigenvalues by thinking geometrically about the corresponding matrix transformation.

1. $\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$.
2. $\begin{bmatrix} -2 & 0 \\ 0 & 4 \end{bmatrix}$.
3. What are the eigenvectors and associated eigenvalues of the identity matrix?
4. What are the eigenvectors and associated eigenvalues of a diagonal matrix with distinct diagonal entries?

6

Suppose that A is a 2×2 matrix having eigenvectors

$$\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

and associated eigenvalues $\lambda_1 = 2$ and $\lambda_2 = -3$. If $\mathbf{x} = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$, find the vector $A^4\mathbf{x}$.

7

Determine whether the following statements are true or false and provide a justification for your response.

1. The eigenvalues of a diagonal matrix are equal to the entries on the diagonal.
2. If $A\mathbf{v} = \lambda\mathbf{v}$, then $A^2\mathbf{v} = \lambda\mathbf{v}$ as well.
3. Every vector is an eigenvector of the identity matrix.
4. If $\lambda = 0$ is an eigenvalue of A, then A is invertible.

5. For every $n \times n$ matrix A, it is possible to find a basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

8

Consider the two special matrices below and find their eigenvectors and associated eigenvalues.

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

9

For each of the following matrix transformations, describe the eigenvalues and eigenvectors of the corresponding matrix A.

1. A reflection in $\mathbb{R}^2$ in the line $y = x$.
2. A 180° rotation in $\mathbb{R}^2$.
3. A 180° rotation in $\mathbb{R}^3$ about the y-axis.
4. A 90° rotation in $\mathbb{R}^3$ about the x-axis.

10

Suppose we have two species, P and Q, where species P preys on Q. Their populations, in millions, in year k are denoted by P_k and Q_k and satisfy

$$\begin{aligned} P_{k+1} &= 0.8P_k + 0.2Q_k \\ Q_{k+1} &= -0.3P_k + 1.5Q_k \end{aligned}.$$

We will keep track of the populations in year k using the vector $\mathbf{x}_k = \begin{pmatrix} P_k \\ Q_k \end{pmatrix}$ so that

$$\mathbf{x}_{k+1} = A\mathbf{x}_k = \begin{bmatrix} 0.8 & 0.2 \\ -0.3 & 1.5 \end{bmatrix} \mathbf{x}_k.$$

1. Show that $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ are eigenvectors of A and find their associated eigenvalues.
2. Suppose that the initial populations are described by the vector $\mathbf{x}_0 = \begin{pmatrix} 38 \\ 44 \end{pmatrix}$. Express $\mathbf{x}_0$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.
3. Find the populations after one year, two years, and three years by writing the vectors $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ as linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$.
4. What is the general form for $\mathbf{x}_k$?
5. After a very long time, what is the ratio of P_k to Q_k?

This page titled 6.1: An introduction to Eigenvalues and Eigenvectors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by David Austin via source content that was edited to the style and standards of the LibreTexts platform.

- 4.1: An introduction to Eigenvalues and Eigenvectors by David Austin is licensed CC BY 4.0. Original source: https://davidaustinm.github.io/ula/ula.html.

6.2: Finding eigenvalues and eigenvectors

The last section introduced eigenvalues and eigenvectors, presented the underlying geometric intuition behind their definition, and demonstrated their use in understanding the long-term behavior of certain systems. We will now develop a more algebraic understanding of eigenvalues and eigenvectors. In particular, we will find an algebraic method for determining the eigenvalues and eigenvectors of a square matrix.

Preview Activity 4.2.1.

Let's begin by reviewing some important ideas that we have seen previously.

1. Suppose that A is a square matrix and that the nonzero vector $\mathbf{x}$ is a solution to the homogeneous equation $A\mathbf{x} = \mathbf{0}$. What can we conclude about the invertibility of A?
2. How does the determinant $\det A$ tell us if there is a nonzero solution to the homogeneous equation $A\mathbf{x} = \mathbf{0}$?
3. Suppose that
$$A = \begin{bmatrix} 3 & -1 & 1 \\ 0 & 2 & 4 \\ 1 & 1 & 3 \end{bmatrix}.$$
Find the determinant $\det A$. What does this tell us about the solution space to the homogeneous equation $A\mathbf{x} = \mathbf{0}$?
4. FInd a basis for $Nul(A)$.
5. What is the relationship between the rank of a matrix and the dimension of its null space?

The characteristic polynomial

We will first see that the eigenvalues of a square matrix appear as the roots of a particular polynomial. To begin, notice that we originally defined an eigenvector as a nonzero vector $\mathbf{v}$ that satisfied the equation $A\mathbf{v} = \lambda\mathbf{v}$. We will rewrite this as

$$\begin{aligned} A\mathbf{v} &= \lambda\mathbf{v} \\ A\mathbf{v} - \lambda\mathbf{v} &= \mathbf{0} \\ A\mathbf{v} - \lambda I\mathbf{v} &= \mathbf{0} \\ (A - \lambda I)\mathbf{v} &= \mathbf{0}. \end{aligned}$$

In other words, an eigenvector $\mathbf{v}$ is a solution of the homogeneous equation $(A - \lambda I)\mathbf{v} = \mathbf{0}$. This puts us in familiar territory, which we will explore in the next activity.

Activity 4.2.2.

The eigenvalues of a square matrix are defined by the condition that there be a nonzero solution to the homogeneous equation $(A - \lambda I)\mathbf{v} = \mathbf{0}$.

1. If there is a nonzero solution to the homogeneous equation $(A - \lambda I)\mathbf{v} = \mathbf{0}$, what can we conclude about the invertibility of the matrix $A - \lambda I$?
2. If there is a nonzero solution to the homogeneous equation $(A - \lambda I)\mathbf{v} = \mathbf{0}$, what can we conclude about the determinant $\det(A - \lambda I)$?
3. Let's consider the matrix
$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$
from which we construct

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

Find the determinant $\det(A - \lambda I)$. What kind of equation do you obtain when we set this determinant to zero to obtain $\det(A - \lambda I) = 0$?

4. Use the determinant you found in the previous part to find the eigenvalues λ by solving $\det(A - \lambda I) = 0$. We considered this matrix in the previous section so we should find the same eigenvalues for A that we found by reasoning geometrically there.
5. Consider the matrix $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ and find its eigenvalues by solving the equation $\det(A - \lambda I) = 0$.
6. Consider the matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and find its eigenvalues by solving the equation $\det(A - \lambda I) = 0$.
7. Find the eigenvalues of the triangular matrix $\begin{bmatrix} 3 & -1 & 4 \\ 0 & -2 & 3 \\ 0 & 0 & 1 \end{bmatrix}$. What is generally true about the eigenvalues of a triangular matrix?

This activity demonstrates a technique that enables us to find the eigenvalues of a square matrix A. Since an eigenvalue λ is a scalar for which the equation $(A - \lambda I)\mathbf{v} = \mathbf{0}$ has a nonzero solution, it must be the case that $A - \lambda I$ is not invertible. Therefore, its determinant is zero. This gives us the equation

$$\det(A - \lambda I) = 0$$

whose solutions are the eigenvalues of A. This equation is called the *characteristic equation* of A.

If we write the characteristic equation for the matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, we see that

$$\det(A - \lambda I) = 0$$

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

$$\begin{aligned} (1-\lambda)^2 - 4 &= 0 \\ -3 - 2\lambda + \lambda^2 &= 0 \\ (3-\lambda)(-1-\lambda) &= 0. \end{aligned}$$

This shows us that the eigenvalues are $\lambda = 3$ and $\lambda = -1$, the same eigenvalues we found by reasoning geometrically in the previous section.

In general, the expression $\det(A - \lambda I)$ is a polynomial in λ, which is called the *characteristic polynomial* of A. If A is an $n \times n$ matrix, the degree of the characteristic polynomial is n. For instance, if A is a 2×2 matrix, then $\det(A - \lambda I)$ is a quadratic polynomial; if A is a 3×3 matrix, then $\det(A - \lambda I)$ is a cubic polynomial.

The other examples that appear in this activity demonstrate some issues we will need to deal with later. For instance, the matrix $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ leads to the characteristic equation, $(2 - \lambda)^2 = 0$. In this case, there is only one eigenvalue $\lambda = 2$ that appears as a repeated root. For now, we simply note that our work in the previous section showed that it was not possible to form a basis of $\mathbb{R}^2$ consisting of eigenvectors of A.

Finally, when $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, we find the characteristic equation $\lambda^2 + 1 = 0$. While this equation has no real solutions, it does have complex solutions $\lambda = \pm i$, and it will be useful for us to work with these complex eigenvalues in the future. In the meantime, remember that this matrix defines a 90° rotation so we do not expect any solutions to the equation $A\mathbf{v} = \lambda\mathbf{v}$ for real eigenvalues λ since a vector $\mathbf{v}$ and $A\mathbf{v}$ can never lie on the same line.

Finally, the eigenvalues of a triangular matrix are easily determined because the determinant of a triangular matrix is the product of the entries on the diagonal. Therefore, the characteristic equation is

$$\det\left(\begin{bmatrix} 3 & -1 & 4 \\ 0 & -2 & 3 \\ 0 & 0 & 1 \end{bmatrix} - \lambda I\right) = \det\begin{bmatrix} 3-\lambda & -1 & 4 \\ 0 & -2-\lambda & 3 \\ 0 & 0 & 1-\lambda \end{bmatrix}$$

$$= (3-\lambda)(-2-\lambda)(1-\lambda) = 0,$$

showing that the eigenvalues are the diagonal entries $\lambda = 3, -2, 1$.

We have how seen now the characteristic equation can be used to determine the eigenvalues of a matrix. Remember, however, that finding the determinant of a matrix using a co-factor expansion is not computationally feasible when the size of the matrix is relatively large. Finding the eigenvalues of a matrix by factoring its characteristic polynomial is therefore a technique limited to relatively small matrices; we will introduce a new technique for finding eigenvalues of larger matrices in the next chapter.

Finding eigenvectors

Now that we can find the eigenvalues of a square matrix A by solving the characteristic equation $\det(A - \lambda I) = 0$, we will turn to the question of finding the eigenvectors associated to an eigenvalue λ. Once again, the key is to note that an eigenvector is a nonzero solution to the homogeneous equation $(A - \lambda I)\mathbf{v} = \mathbf{0}$. In other words, the eigenvectors associated to an eigenvalue λ form the null space $Nul(A - \lambda I)$.

This shows that the eigenvectors associated to an eigenvalue form a subspace of $\mathbb{R}^n$. We will use E_λ to denote the subspace of eigenvectors of a matrix A associated to the eigenvalue λ and note that

$$E_\lambda = Nul(A - \lambda I).$$

We say that E_λ is the *eigenspace* of A associated to the eigenvalue λ.

Activity 4.2.3.

In this activity, we will find the eigenvectors of a matrix as the null space of the matrix $A - \lambda I$.

1. Let's begin with the matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. We have seen that $\lambda = 3$ is an eigenvalue. Form the matrix $A - 3I$ and find a basis for the eigenspace $E_3 = Nul(A - 3I)$. What is the dimension of this eigenspace? For each of the basis vectors $\mathbf{v}$, verify that $A\mathbf{v} = 3\mathbf{v}$.
2. We also saw that $\lambda = -1$ is an eigenvalue. Form the matrix $A - (-1)I$ and find a basis for the eigenspace E_{-1}. What is the dimension of this eigenspace? For each of the basis vectors $\mathbf{v}$, verify that $A\mathbf{v} = -\mathbf{v}$.
3. Is it possible to form a basis of $\mathbb{R}^2$ consisting of eigenvectors of A?
4. Now consider the matrix $A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$. Write the characteristic equation for A and use it to find the eigenvalues of A. For each eigenvalue, find a basis for its eigenspace E_λ. Is it possible to form a basis of $\mathbb{R}^2$ consisting of eigenvectors of A?
5. Next, consider the matrix $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$. Write the characteristic equation for A and use it to find the eigenvalues of A. For each eigenvalue, find a basis for its eigenspace E_λ. Is it possible to form a basis of $\mathbb{R}^2$ consisting of eigenvectors of A?
6. Finally, find the eigenvalues and eigenvectors of the diagonal matrix $A = \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix}$. Explain your result by considering the geometric effect of the matrix transformation defined by A.

Once we find the eigenvalues of a matrix A, describing the eigenspace E_λ amounts to the familiar task of describing the null space $Nul(A - \lambda I)$. For instance, we know that $\lambda = 3$ is an eigenvalue of $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. Then, $E_3 = Nul(A - 3I)$, and we have

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

From the reduced row echelon form, we see that the eigenvectors $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ are determined by the single equation $v_1 - v_2 = 0$ or

$v_1 = v_2$. Therefore the eigenvectors in E_3 have the form

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_2 \\ v_2 \end{pmatrix} = v_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

In other words, E_3 is a one-dimensional subspace of $\mathbb{R}^2$ with basis $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Once again, this agrees with the eigenvectors that we found geometrically in the previous section.

The same reasoning applies to show that the eigenvectors associated to $\lambda = -1$ have the form

$$\mathbf{v} = v_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

which shows that the eigenspace E_{-1} is a one-dimensional subspace of $\mathbb{R}^2$ having basis $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

Two more examples from the activity are important. The characteristic equation for the matrix $A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$ is $\det(A - \lambda I) = (3 - \lambda)^2 = 0$. This shows that there is a single eigenvalue $\lambda = 3$. If we find the eigenspace $E_3 = Nul(A - 3I)$, we have

$$A - 3I = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} - 3I = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

This shows that every vector is in E_3 so that $E_3 = \mathbb{R}^2$. In this case, there is a basis of $\mathbb{R}^2$ consisting of eigenvectors of A. This aligns with our geometric understanding: this matrix has the effect of scaling vectors by a factor of 3 in every direction. Therefore, every vector is an eigenvector with eigenvalue $\lambda = 3$.

However, if we consider the matrix $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$, we find the characteristic equation $(2 - \lambda)^2 = 0$, which shows that there is again a single eigenvalue $\lambda = 2$. In this case,

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

which shows that E_2 is a one-dimensional subspace of $\mathbb{R}^2$ with basis $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Since there are no other eigenvalues, it is not possible to find a basis for $\mathbb{R}^2$ consisting of eigenvectors of A.

Once again, we can understand this result geometrically. The matrix transformation corresponding to the matrix A is a shear that slides vectors horizontally. This transformation therefore only scales vectors that lie on the horizontal axis.

These last two examples illustrate two types of behavior when there is a single eigenvalue. In one case, we are able to construct a basis of $\mathbb{R}^2$ using eigenvectors; in the other, we are not. We will explore this behavior more in the next subsection.

A check on our work.

When finding eigenvalues and their associated eigenvectors in this way, we first find eigenvalues λ by solving the characteristic equation. If λ is a solution to the characteristic equation, then $A - \lambda I$ is not invertible and, consequently, $A - \lambda I$ must contain a row without a pivot position.

This serves as a check on our work. If we row reduce $A - \lambda I$ and find the identity matrix, then we have made an error either in solving the characteristic equation or in finding $Nul(A - \lambda I)$.

The characteristic polynomial and the dimension of eigenspaces

Given a square $n \times n$ matrix A, we saw in the previous section the value of being able to express any vector in $\mathbb{R}^n$ as a linear combination of eigenvectors of A. For this reason, we asked Question 4.1.7 to determine when we can construct a basis of $\mathbb{R}^n$ consisting of eigenvectors. We will explore this question more fully now.

As we saw above, the eigenvalues of A are the solutions of the characteristic equation $\det(A - \lambda I) = 0$. Two examples of characteristic equations we have seen above are

$$(3-\lambda)(-2-\lambda)(1-\lambda) = 0 \quad \text{and} \quad (2-\lambda)^2 = 0.$$

Generally speaking, the characteristic polynomial can always be factored into terms having the form $(\lambda_j - \lambda)$ where λ_j is an eigenvalue of A. In doing so, we must allow ourselves to consider complex eigenvalues, which we will study in more detail in the next section. This means, however, that we can always write the characteristic equation in the form

$$(\lambda_1 - \lambda)^{m_1}(\lambda_2 - \lambda)^{m_2} \ldots (\lambda_p - \lambda)^{m_p} = 0.$$

The solutions to the characteristic equation are the eigenvalues λ_j, and m_j, the number of times that $\lambda_j - \lambda$ appears as a factor in the characteristic polynomial, is called the *multiplicity* of the eigenvalue λ_j.

Example 4.2.1

We have seen that the matrix $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ has the characteristic equation $(2-\lambda)^2 = 0$. This matrix A has a single eigenvalue $\lambda = 2$, which has multiplicity 2.

Example 4.2.2

If a matrix has the characteristic equation

$$(4-\lambda)^2(-5-\lambda)(1-\lambda)^7(3-\lambda)^2 = 0,$$

then that matrix has four eigenvalues: $\lambda = 4$ having multiplicity 2; $\lambda = -5$ having multiplicity 1; $\lambda = 1$ having multiplicty 7; and $\lambda = 3$ having multiplicty 2. The degree of the characteristic polynomial is the sum of the multiplicities $2 + 1 + 7 + 2 = 12$ so this matrix must be a 12×12 matrix.

The multiplicities of the eigenvalues are important because they influence the dimension of the eigenspaces. We know that the dimension of an eigenspace must be at least one; the following proposition also tells us the dimension of an eigenspace can be no larger than the multiplicity of its associated eigenvalue.

Proposition 4.2.3.

If λ is a real eigenvalue of the matrix A with multiplicity m, then

$$1 \leq \dim E_\lambda \leq m.$$

Example 4.2.4

The diagonal matrix $\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$ has the characteristic equation $(3-\lambda)^2 = 0$. There is a single eigenvalue $\lambda = 3$ having multiplicity $m = 2$, and we saw earlier that $\dim E_3 = 2 \leq m = 2$.

Example 4.2.5

The matrix $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ has the characteristic equation $(2-\lambda)^2 = 0$. Once again, there is a single eigenvalue $\lambda = 2$ having multiplicity $m = 2$. In contrast with the previous example, we saw that $\dim E_2 = 1 \leq m = 2$.

Example 4.2.6

We saw earlier that the matrix $\begin{bmatrix} 3 & -1 & 4 \\ 0 & -2 & 3 \\ 0 & 0 & 1 \end{bmatrix}$ has the characteristic equation

$$(3-\lambda)(-2-\lambda)(1-\lambda)=0.$$

There are three eigenvalues $\lambda = 3, -2, 1$ each having multiplicity 1. By the proposition, we are guaranteed that the dimension of each eigenspace is 1; that is,

$$\dim E_3 = \dim E_{-2} = \dim E_1 = 1.$$

It turns out that this is enough to guarantee that there is a basis of $\mathbb{R}^3$ consisting of eigenvectors.

Example 4.2.7

If a 12×12 matrix has the characteristic equation

$$(4-\lambda)^2(-5-\lambda)(1-\lambda)^7(3-\lambda)^2=0,$$

we know there are four eigenvalues $\lambda = 4, -5, 1, 3$. Without more information, all we can say about the dimensions of the eigenspaces is

$$\begin{aligned} 1 \leq \dim E_4 \leq 2 \\ 1 \leq \dim E_{-5} \leq 1 \\ 1 \leq \dim E_1 \leq 7 \\ 1 \leq \dim E_3 \leq 2. \end{aligned}$$

We can guarantee that $\dim E_{-5} = 1$, but we cannot be more specific about the dimensions of the other eigenspaces.

Fortunately, if we have an $n \times n$ matrix, it most commonly happens that the characteristic equation has the form

$$(\lambda_1 - \lambda)(\lambda_2 - \lambda) \ldots (\lambda_n - \lambda) = 0$$

where there are n distinct eigenvalues, each of which has multiplicity 1. In this case, the dimension of each of the eigenspaces $\dim E_{\lambda_j} = 1$. With a little work, it can be seen that choosing a basis vector $\mathbf{v}_j$ for each of the eigenspaces produces a basis for $\mathbb{R}^n$. We therefore have the following proposition.

Proposition 4.2.8.

If A is an $n \times n$ matrix having n distinct real eigenvalues, then there is a basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

This proposition provides one answer to our Question 4.1.7. The next activity explores this question further.

Activity 4.2.4.

1. Identify the eigenvalues, and their multiplicities, of an $n \times n$ matrix whose characteristic polynomial is $(2-\lambda)^3(-3-\lambda)^{10}(5-\lambda)$. What can you conclude about the dimensions of the eigenspaces? What is the dimension of the matrix? Do you have enough information to guarantee that there is a basis of $\mathbb{R}^n$ consisting of eigenvectors?
2. Find the eigenvalues of $\begin{bmatrix} 0 & -1 \\ 4 & -4 \end{bmatrix}$ and state their multiplicities. Can you find a basis of $\mathbb{R}^2$ consisting of eigenvectors of this matrix?

3. Consider the matrix $A = \begin{bmatrix} -1 & 0 & 2 \\ -2 & -2 & -4 \\ 0 & 0 & -2 \end{bmatrix}$ whose characteristic equation is

$$(-2-\lambda)^2(-1-\lambda) = 0.$$

   1. Identify the eigenvalues and their multiplicities.
   2. For each eigenvalue λ, find a basis of the eigenspace E_λ and state its dimension.
   3. Is there a basis of $\mathbb{R}^3$ consisting of eigenvectors of A?

4. Now consider the matrix $A = \begin{bmatrix} -5 & -2 & -6 \\ -2 & -2 & -4 \\ 2 & 1 & 2 \end{bmatrix}$ whose characteristic equation is also

$$(-2-\lambda)^2(-1-\lambda) = 0.$$

   1. Identify the eigenvalues and their multiplicities.
   2. For each eigenvalue λ, find a basis of the eigenspace E_λ and state its dimension.
   3. Is there a basis of $\mathbb{R}^3$ consisting of eigenvectors of A?

5. Consider the matrix $A = \begin{bmatrix} -5 & -2 & -6 \\ 4 & 1 & 8 \\ 2 & 1 & 2 \end{bmatrix}$ whose characteristic equation is

$$(-2-\lambda)(1-\lambda)(-1-\lambda) = 0.$$

   1. Identify the eigenvalues and their multiplicities.
   2. For each eigenvalue λ, find a basis of the eigenspace E_λ and state its dimension.
   3. Is there a basis of $\mathbb{R}^3$ consisting of eigenvectors of A?

Computational issues in finding eigenvalues and eigenvectors

We can use Sage to find the characteristic polynomial, eigenvalues, and eigenvectors of a matrix. As we will see, however, some care is required when dealing with matrices whose entries include floating point numbers. The next activity demonstrates how Sage can be used in this way and some of the complications that arise. We will revisit this issue in subsequent sections.

Activity 4.2.5.

We will use Sage to find the eigenvalues and eigenvectors of a matrix. Let's begin with the matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$.

1. We can find the characteristic polynomial of a matrix A by writing `A.charpoly('lam')`. Notice that we have to give Sage a variable in which to write the polynomial; here, we use `lam` though you could just as well use `x`.

   The factored form of the characteristic polynomial may be more useful since it will tell us the eigenvalues and their multiplicities. The factor chacteristic polynomial is found with `A.fcp('lam')`.

2. If we only want the eigenvalues, we can use `A.eigenvalues()`.

   Notice that the multiplicity of an eigenvalue is the number of times it is repeated in the list of eigenvalues.

3. Finally, we can find eigenvectors by `A.eigenvectors_right()`. (We are looking for *right* eigenvalues since the vector $\mathbf{v}$ appears to the right of A in the definition $A\mathbf{v} = \lambda\mathbf{v}$.)

   At first glance, the result of this command can be a little confusing to interpret. What we see is a list with one entry for each eigenvalue. For each eigenvalue, there is a triple consisting of (i) the eigenvalue λ, (ii) a basis for E_λ, and (iii) the multiplicity of λ.

4. When working with decimal entries, which are called *floating point numbers* in computer science, we must remember that computers perform only approximate arithmetic. This is a problem when we wish to find the eigenvectors of such a matrix. To illustrate, consider the matrix $A = \begin{bmatrix} 0.4 & 0.3 \\ 0.6 & 0.7 \end{bmatrix}$.

1. Without using Sage, find the eigenvalues of this matrix.
2. What do you find for the reduced row echelon form of $A - I$?
3. Let's now use Sage to determine the reduced row echelon form of $A - I$:

   What result does Sage report for the reduced row echelon form? Why is this result not correct?

4. Because the arithmetic Sage performs with floating point entries is only approximate, we are not able to find the eigenspace E_1. In this next chapter, we will learn how to address this issue. In the meantime, we can get around this problem by writing the entries in the matrix as rational numbers:

Summary

In this section, we developed a technique for finding the eigenvalues and eigenvectors of an $n \times n$ matrix A.

- The expression $\det(A - \lambda I)$ is a degree n polynomial, known as the characteristic polynomial. The eigenvalues are the roots of the characteristic polynomial $\det(A - \lambda I) = 0$.
- The set of eigenvectors associated to the eigenvalue λ forms the eigenspace $E_\lambda = Nul(A - \lambda I)$.
- If the factor $(\lambda_j - \lambda)$ appears m_j times in the characteristic polynomial, we say that the eigenvalue λ_j has multiplicity m_j and note that

$$1 \leq \dim E_{\lambda_j} \leq m_j.$$

- If each of the eigenvalues is real and has multiplicity 1, then we can form a basis for $\mathbb{R}^n$ consisting of eigenvectors of A.
- We can use Sage to find the eigenvalues and eigenvalues of matrices. However, we need to be careful working with floating point numbers since floating point arithmetic is only an approximation.

Exercises 4.2.6Exercises

1

For each of the following matrices, find its characteristic polynomial, its eigenvalues, and the multiplicity of each eigenvalue.

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

2

Given an $n \times n$ matrix A, an important question Question 4.1.7 asks whether we can find a basis of $\mathbb{R}^n$ consisting of eigenvectors of A. For each of the matrices in the previous exericse, find a basis of $\mathbb{R}^n$ consisting of eigenvectors or state why such a basis does not exist.

3

Determine whether the following statements are true or false and provide a justification for your response.

1. The eigenvalues of a matrix A are the entries on the diagonal of A.
2. If λ is an eigenvalue of multiplicity 1, then E_λ is one-dimensional.

3. If a matrix A is invertible, then $\lambda = 0$ cannot be an eigenvalue.
4. If A is a 13×13 matrix, the charasteristic polynomial has degree less than 13.
5. The eigenspace E_λ of A is the same as the null space $Nul(A - \lambda I)$.

4

Provide a justification for your response to the following questions.

1. Suppose that A is a 3×3 matrix having eigenvalues $\lambda = -3, 3, -5$. What are the eigenvalues of $2A$?
2. Suppose that D is a diagonal 3×3 matrix. Why can you guarantee that there is a basis of $\mathbb{R}^3$ consisting of eigenvectors of D?
3. If A is a 3×3 matrix whose eigenvalues are $\lambda = -1, 3, 5$, can you guarantee that there is a basis of $\mathbb{R}^3$ consisting of eigenvectors of A?
4. Suppose that the characteristic polynomial of a matrix A is
$$\det(A - \lambda I) = -\lambda^3 + 4\lambda.$$
What are the eigenvalues of A? Is A invertible? Is there a basis of $\mathbb{R}^n$ consisting of eigenvectors of A?
5. If the characteristic polynomial of A is
$$\det(A - \lambda I) = (4 - \lambda)(-2 - \lambda)(1 - \lambda),$$
what is the characteristic polynomial of A^2? what is the characteristic polynomial of A^{-1}?

5

For each of the following matrices, use Sage to determine its eigenvalues, their multiplicities, and a basis for each eigenspace. For which matrices is it possible to construct a basis for $\mathbb{R}^3$ consisting of eigenvectors?

1. $A = \begin{bmatrix} -4 & 12 & -6 \\ 4 & -5 & 4 \\ 11 & -20 & 13 \end{bmatrix}$
2. $A = \begin{bmatrix} 1 & -3 & 1 \\ -4 & 8 & -5 \\ -8 & 17 & -10 \end{bmatrix}$
3. $A = \begin{bmatrix} 3 & -8 & 4 \\ -2 & 3 & -2 \\ -6 & 12 & -7 \end{bmatrix}$

6

There is a relationship between the determinant of a matrix and the product of its eigenvalues.

1. We have seen that the eigenvalues of the matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ are $\lambda = 3, -1$. What is $\det A$? What is the product of the eigenvalues of A?
2. Consider the triangular matrix $A = \begin{bmatrix} 2 & 0 & 0 \\ -1 & -3 & 0 \\ 3 & 1 & -2 \end{bmatrix}$. What are the eigenvalues of A? What is $\det A$? What is the product of the eigenvalues of A?
3. Based on these examples, what do you think is the relationship between the determinant of a matrix and the product of its eigenvalues?
4. Suppose the characteristic polynomial is written as

$$\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \ldots (\lambda_n - \lambda).$$

By substituting $\lambda = 0$ into this equation, explain why the determinant of a matrix equals the product of its eigenvalues.

7

Consider the matrix $A = \begin{bmatrix} 0.5 & 0.6 \\ -0.3 & 1.4 \end{bmatrix}$.

1. Find the eigenvalues of A and a basis for their associated eigenspaces.
2. Suppose that $\mathbf{x}_0 = \begin{pmatrix} 11 \\ 6 \end{pmatrix}$. Express $\mathbf{x}_0$ as a linear combination of eigenvectors of A.
3. Define the vectors

$$\begin{aligned} \mathbf{x}_1 &= A\mathbf{x}_0 \\ \mathbf{x}_2 &= A\mathbf{x}_1 = A^2\mathbf{x}_0 \\ \mathbf{x}_3 &= A\mathbf{x}_2 = A^3\mathbf{x}_0 \\ &\vdots = \vdots \end{aligned}.$$

   Write $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ as a linear combination of eigenvectors of A.

4. What happens to $\mathbf{x}_k$ as k grows larger and larger?

8

Consider the matrix $A = \begin{bmatrix} 0.4 & 0.3 \\ 0.6 & 0.7 \end{bmatrix}$

1. Find the eigenvalues of A and a basis for their associated eigenspaces.
2. Suppose that $\mathbf{x}_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Express $\mathbf{x}_0$ as a linear combination of eigenvectors of A.
3. Define the vectors

$$\begin{aligned} \mathbf{x}_1 &= A\mathbf{x}_0 \\ \mathbf{x}_2 &= A\mathbf{x}_1 = A^2\mathbf{x}_0 \\ \mathbf{x}_3 &= A\mathbf{x}_2 = A^3\mathbf{x}_0 \\ &\vdots = \vdots \end{aligned}.$$

   Write $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ as a linear combination of eigenvectors of A.

4. What happens to $\mathbf{x}_k$ as k grows larger and larger?

This page titled 6.2: Finding eigenvalues and eigenvectors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by David Austin via source content that was edited to the style and standards of the LibreTexts platform.

- 4.2: Finding eigenvalues and eigenvectors by David Austin is licensed CC BY 4.0. Original source: https://davidaustinm.github.io/ula/ula.html.

6.3: Diagonalization, similarity, and powers of a matrix

The first example we considered in this chapter was the matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, which has eigenvectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ and associated eigenvalues $\lambda_1 = 3$ and $\lambda_2 = -1$. In Subsection 4.1.2, we described how A is, in some sense, equivalent to the diagonal matrix $D = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}$.

This equivalence is summarized by Figure 4.3.1. The diagonal matrix D has the geometric effect of stretching vectors horizontally by a factor of 3 and flipping vectors vertically. The matrix A has the geometric effect of stretching vectors by a factor of 3 in the direction $\mathbf{v}_1$ and flipping them in the direction of $\mathbf{v}_2$. The geometric effect of A is the same as that of D when viewed in a basis of eigenvectors of A.

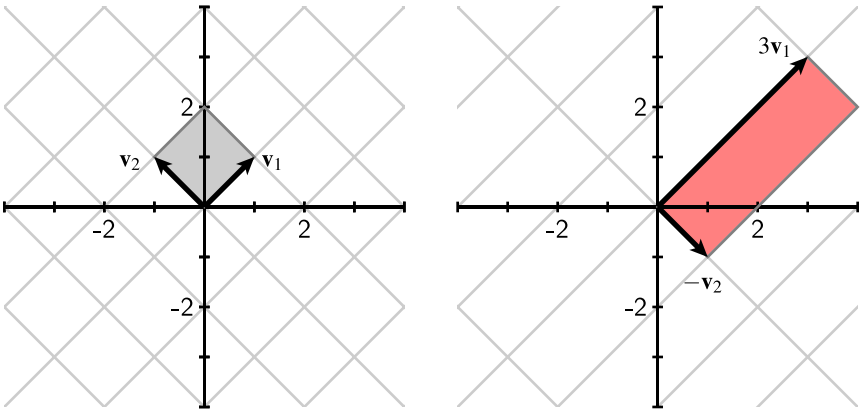


Figure 6.3.1: The matrix A has the same geometric effect as the diagonal matrix D when expressed in the coordinate system defined by the basis of eigenvectors.

Now that we have developed some algebraic techniques for finding eigenvalues and eigenvectors, we will explore this observation more deeply. In particular, we will make precise the sense in which A and D are equivalent by using the coordinate system defined by the basis of eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

Preview Activity 4.3.1.

Let's recall how a vector in $\mathbb{R}^2$ can be represented in a coordinate system defined by a basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$.

1. Suppose that we consider the basis $\mathcal{B}$ defined by

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}. \tag{6.3.1}$$

   Find the vector $\mathbf{x}$ whose representation in the coordinate system defined by $\mathcal{B}$ is $\{\mathbf{x}\}_{\mathcal{B}} = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$.

2. Consider the vector $\mathbf{x} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$ and find its representation $\{\mathbf{x}\}_{\mathcal{B}}$ in the coordinate system defined by $\mathcal{B}$.
3. How do we use the matrix $C_{\mathcal{B}} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}$ to convert a vector's representation $\{\mathbf{x}\}_{\mathcal{B}}$ in the coordinate system defined by $\mathcal{B}$ into its standard representation $\mathbf{x}$? How do we use this matrix to convert $\mathbf{x}$ into $\{\mathbf{x}\}_{\mathcal{B}}$?
4. Suppose that we have a matrix A whose eigenvectors are $\mathbf{v}_1$ and $\mathbf{v}_2$ and associated eigenvalues are $\lambda_1 = 4$ and $\lambda_2 = 2$. Express the vector $A(-3\mathbf{v}_1 + 5\mathbf{v}_2)$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.
5. If $\{\mathbf{x}\}_{\mathcal{B}} = \begin{pmatrix} -3 \\ 5 \end{pmatrix}$, find $\{A\mathbf{x}\}_{\mathcal{B}}$.

Diagonalization of matrices

As we have investigated eigenvalues and eigenvectors of matrices in this chapter, we have frequently asked whether we can find a basis of eigenvectors, as in Question 4.1.7. In fact, Proposition 4.2.3 tells us that if A is an $n \times n$ matrix having distinct and real eigenvalues, then there is a basis for $\mathbb{R}^n$ consisting of eigenvectors of A. There are, in addition, other conditions on A that guarantee such a basis, as we will see in subsequent chapters, but for now, suffice it to say that for many matrices, we can find a basis of eigenvectors. We will now see how such a matrix A is equivalent to a diagonal matrix D.

Remember also that we have seen how to use a basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of $\mathbb{R}^n$ to construct a coordinate system for $\mathbb{R}^n$. In particular, $\{\mathbf{x}\}_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$ if $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n$. We also used matrix multiplication to express this fact: if $C_{\mathcal{B}} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{bmatrix}$, then

$$\mathbf{x} = C_{\mathcal{B}}\{\mathbf{x}\}_{\mathcal{B}}, \qquad \{\mathbf{x}\}_{\mathcal{B}} = C_{\mathcal{B}}^{-1}\mathbf{x}. \tag{6.3.2}$$

Activity 4.3.2.

Once again, we will consider the matrices

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \qquad D = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}. \tag{6.3.3}$$

The matrix A has eigenvectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ and eigenvalues $\lambda_1 = 3$ and $\lambda_2 = -1$. We will consider the basis of $\mathbb{R}^2$ consisting of eigenvectors $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$.

1. If $\mathbf{x} = 2\mathbf{v}_1 - 3\mathbf{v}_2$, write $A\mathbf{x}$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.
2. If $\{\mathbf{x}\}_{\mathcal{B}} = \begin{pmatrix} 2 \\ -3 \end{pmatrix}$, find $\{A\mathbf{x}\}_{\mathcal{B}}$, the representation of $A\mathbf{x}$ in the coordinate system defined by $\mathcal{B}$.
3. If $\{\mathbf{x}\}_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$, find $\{A\mathbf{x}\}_{\mathcal{B}}$, the representation of $A\mathbf{x}$ in the coordinate system defined by $\mathcal{B}$.
4. Explain why $\{A\mathbf{x}\}_{\mathcal{B}} = D\{\mathbf{x}\}_{\mathcal{B}}$.
5. Explain why $C_{\mathcal{B}}^{-1}A\mathbf{x} = DC_{\mathcal{B}}^{-1}\mathbf{x}$ for all vectors $\mathbf{x}$ and hence
$$C_{\mathcal{B}}^{-1}A = DC_{\mathcal{B}}^{-1}. \tag{6.3.4}$$
6. Explain why $A = C_{\mathcal{B}}DC_{\mathcal{B}}^{-1}$ and verify this relationship by computing $C_{\mathcal{B}}DC_{\mathcal{B}}^{-1}$ in the Sage cell below.

The key to understanding the equivalence of a matrix A and a diagonal matrix D is through the coordinate system defined by a basis consisting of eigenvectors of A. We will assume that A is an $n \times n$ matrix and that there is a basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ consisting of eigenvectors of A with associated eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.

We know that if

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n, \tag{6.3.5}$$

then

$$A\mathbf{x} = \lambda_1c_1\mathbf{v}_1 + \lambda_2c_2\mathbf{v}_2 + \ldots + \lambda_nc_n\mathbf{v}_n. \tag{6.3.6}$$

This fact is conveniently expressed using the coordinate system defined by $\mathcal{B}$; in particular,

$$\{\mathbf{x}\}_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}, \qquad \{A\mathbf{x}\}_{\mathcal{B}} = \begin{pmatrix} \lambda_1c_1 \\ \lambda_2c_2 \\ \vdots \\ \lambda_nc_n \end{pmatrix}. \tag{6.3.7}$$

Forming the diagonal matrix

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}, \tag{6.3.8}$$

we see that

$$\{A\mathbf{x}\}_{\mathcal{B}} = D\{\mathbf{x}\}_{\mathcal{B}}. \tag{6.3.9}$$

We now use the fact that the matrix $C_{\mathcal{B}} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{bmatrix}$ performs the change of coordinates; that is, $\{A\mathbf{x}\}_{\mathcal{B}} = C_{\mathcal{B}}^{-1}A\mathbf{x}$ and $\{\mathbf{x}\}_{\mathcal{B}} = C_{\mathcal{B}}^{-1}\mathbf{x}$. This says that

$$C_{\mathcal{B}}^{-1}A\mathbf{x} = DC_{\mathcal{B}}^{-1}\mathbf{x}, \tag{6.3.10}$$

for all vectors $\mathbf{x}$, which means that $C_{\mathcal{B}}^{-1}A = DC_{\mathcal{B}}^{-1}$ or

$$A = C_{\mathcal{B}}DC_{\mathcal{B}}^{-1}. \tag{6.3.11}$$

So that the form of this expression stands out more clearly, it is customary to denote the matrix $C_{\mathcal{B}}$ as P so that we have $P = C_{\mathcal{B}}$ and hence

$$A = PDP^{-1}. \tag{6.3.12}$$

Definition 4.3.2

We say that the matrix A is *diagonalizable* if there is a diagonal matrix D and invertible matrix P such that

$$A = PDP^{-1}. \tag{6.3.13}$$

This is the sense in which we mean that A is equivalent to a diagonal matrix D. The expression $A = PDP^{-1}$ says that A, expressed in the basis defined by the columns of P, has the same geometric effect as D, expressed in the standard basis $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$.

We have now seen the following proposition.

Proposition 4.3.3.

If A is an $n \times n$ matrix and there is a basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ consisting of eigenvectors of A having associated eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, then A is diagonalizable. That is, we can write $A = PDP^{-1}$ where D is the diagonal matrix whose diagonal entries are the eigenvalues of A

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \tag{6.3.14}$$

and the matrix $P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{bmatrix}$.

In fact, if we only know that $A = PDP^{-1}$ where $P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots \mathbf{v}_n \end{bmatrix}$, we can say that the vectors $\mathbf{v}_j$ are eigenvectors of A and that the associated eigenvalue is the jth diagonal entry of D.

Example 4.3.4

We will try to find a diagonalization of $A = \begin{bmatrix} -5 & 6 \\ -3 & 4 \end{bmatrix}$.

First, we find the eigenvalues of A by solving the characteristic equation

$$\det(A - \lambda I) = (-5 - \lambda)(4 - \lambda) + 18 = (-2 - \lambda)(1 - \lambda) = 0. \tag{6.3.15}$$

This shows that the eigenvalues of A are $\lambda_1 = -2$ and $\lambda_2 = 1$.

By constructing $Nul(A - (-2)I)$, we find a basis for E_{-2} consisting of the vector $\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Similarly, a basis for E_1 consists of the vector $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. This shows that we can construct a basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ of $\mathbb{R}^2$ consisting of eigenvectors of A.

We now form the matrices

$$D = \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix}, \qquad P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \tag{6.3.16}$$

and verify that

$$PDP^{-1} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} -5 & 6 \\ -3 & 4 \end{bmatrix} = A. \tag{6.3.17}$$

There are, of course, many ways to diagonalize A. For instance, we could change the order of the eigenvalues and eigenvectors and write

$$D = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}, \qquad P = \begin{bmatrix} \mathbf{v}_2 & \mathbf{v}_1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}. \tag{6.3.18}$$

If we choose a different basis for the eigenspaces, we will also find a different matrix P that diagonalizes A. The point is that there are many ways in which A can be written in the form $A = PDP^{-1}$.

Example 4.3.5

We will try to find a diagonalization of $A = \begin{bmatrix} 0 & 4 \\ -1 & 4 \end{bmatrix}$.

Once again, we find the eigenvalues by solving the characteristic equation:

$$\det(A - \lambda I) = -\lambda(4 - \lambda) + 4 = (2 - \lambda)^2 = 0. \tag{6.3.19}$$

In this case, there is a single eigenvalue $\lambda = 2$.

We find a basis for the eigenspace E_2 by describing $Nul(A - 2I)$:

$$A - 2I = \begin{bmatrix} -2 & 4 \\ -1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix}. \tag{6.3.20}$$

This shows that the eigenspace E_2 is one-dimensional with $\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ forming a basis.

In this case, there is not a basis of $\mathbb{R}^2$ consisting of eigenvectors of A, which tells us that A is not diagonalizable.

Example 4.3.6

Suppose we know that $A = PDP^{-1}$ where

$$D = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}, \qquad P = \begin{bmatrix} \mathbf{v}_2 & \mathbf{v}_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}. \tag{6.3.21}$$

In this case, we know that the columns of P form eigenvectors of A. For instance, $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector of A with eigenvalue $\lambda_1 = 2$. Also, $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is an eigenvector with eigenvalue $\lambda_2 = -2$.

We can verify this by computing

$$A = PDP^{-1} = \begin{bmatrix} 6 & -4 \\ 8 & -6 \end{bmatrix}. \tag{6.3.22}$$

Then, we can compute that $A\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2\mathbf{v}_1$ and $A\mathbf{v}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} = -2\mathbf{v}_2$.

Activity 4.3.3.

1. Find a diagonalization of A, if one exists, when

$$A = \begin{bmatrix} 3 & -2 \\ 6 & -5 \end{bmatrix}. \tag{6.3.23}$$

2. Can the diagonal matrix

$$A = \begin{bmatrix} 2 & 0 \\ 0 & -5 \end{bmatrix} \tag{6.3.24}$$

be diagonalized? If so, explain how to find the matrices P and D.

3. Find a diagonalization of A, if one exists, when

$$A = \begin{bmatrix} -2 & 0 & 0 \\ 1 & -3 & 0 \\ 2 & 0 & -3 \end{bmatrix}. \tag{6.3.25}$$

4. Find a diagonalization of A, if one exists, when

$$A = \begin{bmatrix} -2 & 0 & 0 \\ 1 & -3 & 0 \\ 2 & 1 & -3 \end{bmatrix}. \tag{6.3.26}$$

5. Suppose that $A = PDP^{-1}$ where

$$D = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}, \qquad P = \begin{bmatrix} \mathbf{v}_2 & \mathbf{v}_1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix}. \tag{6.3.27}$$

   1. Explain why A is invertible.
   2. Find a diagonalization of A^{-1}.
   3. Find a diagonalization of A^3.

Powers of a diagonalizable matrix

In several earlier examples, we have been interested in computing powers of a given matrix. For instance, in Activity 4.1.3, we are given the matrix $A = \begin{bmatrix} 0.8 & 0.6 \\ 0.2 & 0.4 \end{bmatrix}$ and an initial vector $\mathbf{x}_0 = \begin{pmatrix} 1000 \\ 0 \end{pmatrix}$, and we wanted to compute

$$\begin{aligned} \mathbf{x}_1 &= A\mathbf{x}_0 \\ \mathbf{x}_2 &= A\mathbf{x}_1 = A^2\mathbf{x}_0 \\ \mathbf{x}_3 &= A\mathbf{x}_2 = A^3\mathbf{x}_0. \end{aligned} \tag{6.3.28}$$

More generally, we would like to find $\mathbf{x}_k = A^k\mathbf{x}_0$ and determine what happens as k becomes very large. If a matrix A is diagonalizable, writing $A = PDP^{-1}$ can help us understand powers of A easily.

Activity 4.3.4.

1. Let's begin with the diagonal matrix

$$D = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}. \tag{6.3.29}$$

   Find the powers D^2, D^3, and D^4. What is D^k for a general value of k?

2. Suppose that A is a matrix with eigenvector $\mathbf{v}$ and associated eigenvalue λ; that is, $A\mathbf{v} = \lambda\mathbf{v}$. By considering $A^2\mathbf{v}$, explain why $\mathbf{v}$ is also an eigenvector of A with eigenvalue λ^2.
3. Suppose that $A = PDP^{-1}$ where

$$D = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}. \tag{6.3.30}$$

   Remembering that the columns of P are eigenvectors of A, explain why A^2 is diagonalizable and find a diagonalization of it.

4. Give another explanation of the diagonalizability of A^2 by writing

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD(P^{-1}P)DP^{-1}. \tag{6.3.31}$$

5. In the same way, find a diagonalization of A^3, A^4, and A^k.
6. Suppose that A is a diagonalizable 2×2 matrix with eigenvalues $\lambda_1 = 0.5$ and $\lambda_2 = 0.1$. What happens to A^k as k becomes very large?

We begin by noting that the eigenvectors of a matrix A are also eigenvectors of the powers of A. For instance, if $A\mathbf{v} = \lambda\mathbf{v}$, then

$$A^2\mathbf{v} = A(A\mathbf{v}) = A(\lambda\mathbf{v}) = \lambda A\mathbf{v} = \lambda^2\mathbf{v}. \tag{6.3.32}$$

In this way, we see that $\mathbf{v}$ is an eigenvector of A^2 with eigenvalue λ^2. Furthermore, for any k, $\mathbf{v}$ is an eigenvector of A^k with eigenvalue λ^k.

Now if A is diagonalizable, we can write $A = PDP^{-1}$ where the columns of P are eigenvectors of A and the diagonal entries of D are the eigenvalues. If $D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$, then

$$A^2 = P\begin{bmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{bmatrix}P^{-1} = PD^2P^{-1}. \tag{6.3.33}$$

We have the same matrix P in this expression since the eigenvectors of A^2 are also the eigenvectors of A.

Another way to see this is to note that

$$\begin{aligned} A^2 &= (PDP^{-1})(PDP^{-1}) \\ &= PD(P^{-1}P)DP^{-1} \\ &= PDIDP^{-1} \\ &= PDDP^{-1} \\ &= PD^2P^{-1}. \end{aligned} \tag{6.3.34}$$

Similarly, any power of A is diagaonalizable; in particular, $A^k = PD^kP^{-1}$.

In the next section, we will see some important uses of our ability to deal with powers in this way. Until then, consider the case where $D = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.1 \end{bmatrix}$ so that $D^k = \begin{bmatrix} 0.5^k & 0 \\ 0 & 0.1^k \end{bmatrix}$. As k becomes very large, the diagonal entries become increasingly close to zero. This means that D^k becomes increasingly close to the zero matrix $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ as does $A^k = PD^kP^{-1}$. In other words, no matter what vector $\mathbf{x}_0$ we begin with, the vectors $A^k\mathbf{x}_0$ becomes increasingly close to $\mathbf{0}$.

Similarity and complex eigenvalues

We have been interested in diagonalizing a matrix A because doing so relates a matrix A to a simpler diagonal matrix D. If we write $A = PDP^{-1}$, we see that multiplying a vector by A in the coordinates defined by the columns of P is the same as multiplying by D in standard coordinates. Under this change of coordinates, A and D have the same effect on vectors.

More generally, if we have two matrices A and B such that $A = PBP^{-1}$, we may regard multiplication by A and B as having the same effect on vectors under the change of coordinates defined by the columns of P. That is, if $\mathcal{B}$ is the basis formed by the columns of P, then $\{A\mathbf{x}\}_{\mathcal{B}} = B\{\mathbf{x}\}_{\mathcal{B}}$. This leads to the following definition.

Definition 4.3.7

We say that A is *similar* to B if there is an invertible matrix P such that $A = PBP^{-1}$.

Notice that a matrix is diagonalizable if and only if it is similar to a diagonal matrix. We have, however, seen several examples of a matrix A that is not diagonalizable. In this case, it is natural to ask if there is some simpler matrix that is similar to A.

Example 4.3.8

Let's consider the matrix $A = \begin{bmatrix} -2 & 2 \\ -5 & 4 \end{bmatrix}$ whose characteristic equation is

$$\det(A - \lambda I) = (-2 - \lambda)(4 - \lambda) + 10 = 2 - 2\lambda + \lambda^2 = 0. \tag{6.3.35}$$

Applying the quadratic formula to find the eigenvalues, we obtain

$$\lambda = \frac{2 \pm \sqrt{(-2)^2 - 4 \cdot 1 \cdot 2}}{2} = 1 \pm i. \tag{6.3.36}$$

Here we see that the matrix A has two complex eigenvalues and is therefore not diagonalizable.

In case a matrix A has complex eigenvalues, we will find a simpler matrix C that is similar to A. In particular, if A has an eigenvalue $\lambda = a + bi$, then A is similar to $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$.

The next activity shows that C has a simple geometric effect on $\mathbb{R}^2$. First, however, we will rewrite C in polar coordinates, as shown in the figure. We form the point (a, b), which defines r, the distance from the origin, and θ, the angle formed with the positive horizontal axis. We then have

$$\begin{aligned} a &= r\cos\theta \\ b &= r\sin\theta. \end{aligned} \tag{6.3.37}$$

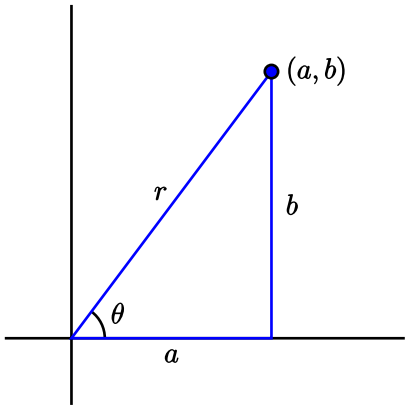


Figure 6.3.2

Notice that $r = \sqrt{a^2 + b^2}$.

Activity 4.3.5.

1. We will rewrite C in terms of r and θ. Explain why

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} r\cos\theta & -r\sin\theta \\ r\sin\theta & r\cos\theta \end{bmatrix} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{6.3.38}$$

2. Explain why C has the geometric effect of rotating vectors by θ and stretching them by a factor of r.
3. Let's now consider the matrix A from Example 4.3.8:

$$A = \begin{bmatrix} -2 & 2 \\ -5 & 4 \end{bmatrix} \tag{6.3.39}$$

   whose eigenvalues are $\lambda_1 = 1 + i$ and $\lambda_2 = 1 - i$. We will choose to focus on one of the eigenvalues $\lambda_1 = a + bi = 1 + i$.

   Form the matrix C using these values of a and b. Then rewrite the point (a, b) in polar coordinates by identifying the values of r and θ. Explain the geometric effect of multiplying vectors of C.

4. Suppose that $P = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$. Verify that $A = PCP^{-1}$.
5. Explain why $A^k = PC^kP^{-1}$.
6. We formed the matrix C by choosing the eigenvalue $\lambda_1 = 1 + i$. Suppose we had instead chosen $\lambda_2 = 1 - i$. Form the matrix C' and use polar coordinates to describe the geometric effect of C.
7. Using the matrix $P' = \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix}$, show that $A = P'C'P'^{-1}$.

If the 2×2 matrix A has a complex eigenvalue $\lambda = a + bi$, this activity demonstrates the fact that A is similar to the matrix $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. When we consider the matrix $A = \begin{bmatrix} -2 & 2 \\ -5 & 4 \end{bmatrix}$, we find the complex eigenvalue $\lambda = 1 + i$, which leads to the matrix

$$C = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \cos(45^\circ) & -\sin(45^\circ) \\ \sin(45^\circ) & \cos(45^\circ) \end{bmatrix}. \tag{6.3.40}$$

The matrix has the geometric effect of rotating vectors by 45° and stretching them by a factor of $\sqrt{2}$, as shown in the figure.

As we saw in the activity, our original matrix A is similar to C. That is, we saw that there is a matrix P such that $A = PCP^{-1}$. This

means that, when expressed in the coordinates defined by the columns of P, multiplying a vector by A is equivalent to multiplying by C; that is, if $\mathcal{B}$ is the basis formed by the columns of A, then $\{A\mathbf{x}\}_{\mathcal{B}} = C\{\mathbf{x}\}_{\mathcal{B}}$.

Had we chosen the other eigenvalue $\lambda_2 = 1 - i$, we would have formed the matrix

$$C' = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \cos(-45^\circ) & -\sin(-45^\circ) \\ \sin(-45^\circ) & \cos(-45^\circ) \end{bmatrix}. \tag{6.3.41}$$

In other words, this matrix C' rotates vectors by -45° and stretches them by a factor of $\sqrt{2}$. The original matrix A is also similar to C'.

Depending on which complex eigenvalue we choose, we find a matrix C that performs either a counterclockwise or a clockwise rotation. In our future uses, we will focus on r, the streching factor, and not be concerned about the direction of the rotation.

Summary

The ideas in this section demonstrate how the eigenvalues and eigenvectors of a matrix A can provide us with a new coordinate system in which multiplying by A reduces to a simpler operation.

- We said that A is diagonalizable if we can write $A = PDP^{-1}$ where D is a diagonal matrix. The columns of P consist of eigenvectors of A and the diagonal entries of D are the associated eigenvalues.
- An $n \times n$ matrix A is diagonalizable if and only if there is a basis of $\mathbb{R}^n$ consisting of eigenvectors of A.
- We said that A and B are similar if there is an invertible matrix P such that $A = PBP^{-1}$. In this case, $A^k = PB^kP^{-1}$.
- If A is a 2×2 matrix with complex eigenvalue $\lambda = a + bi$, then A is similar to $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. Writing the point (a, b) in polar coordinates r and θ, we see that C rotates vectors through an angle θ and stretches them by a factor of $r = \sqrt{a^2 + b^2}$.

Exercises 4.3.5Exercises

1

Determine whether the following matrices are diagonalizable. If so, find matrices D and P such that $A = PDP^{-1}$.

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

2

Determine whether the following matrices have complex eigenvalues. If so, find the matrix C such that $A = PCP^{-1}$.

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

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

3

Determine whether the following statements are true or false and provide a justification for your response.

1. If A is invertible, then A is diagonalizable.
2. If A and B are similar and A is invertible, then B is also invertible.
3. If A is a diagonalizable $n \times n$ matrix, then there is a basis of $\mathbb{R}^n$ consisting of eigenvectors of A.
4. If A is diagonalizable, then A^{10} is also diagonalizable.
5. If A is diagonalizable, then A is invertible.

4

Provide a justification for your response to the following questions.

1. If A is a 3×3 matrix having eigenvalues $\lambda = 2, 3, -4$, can you guarantee that A is diagonalizable?
2. If A is a 2×2 matrix with a complex eigenvalue, can you guarantee that A is diagonalizable?
3. If A is similar to the matrix $B = \begin{bmatrix} -5 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 3 \end{bmatrix}$, is A diagonalizable?
4. What matrices are similar to the identity matrix?
5. If A is a diagonalizable 2×2 matrix with a single eigenvalue $\lambda = 4$, what is A?

5

Describe geometric effect that the following matrices have on $\mathbb{R}^2$:

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

6

We say that A is similar to B if there is a matrix P such that $A = PBP^{-1}$.

1. If A is similar to B, explain why B is similar to A.
2. If A is similar to B and B is similar to C, explain why A is similar to C.
3. If A is similar to B and B is diagonalizable, explain why A is diagonalizable.
4. If A and B are similar, explain why A and B have the same characteristic polynomial; that is, explain why $\det(A - \lambda I) = \det(B - \lambda I)$.
5. If A and B are similar, explain why A and B have the same eigenvalues.

7

Suppose that $A = PDP^{-1}$ where

$$D = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad P = \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}. \tag{6.3.42}$$

1. Explain the geometric effect that D has on vectors in $\mathbb{R}^2$.
2. Explain the geometric effect that A has on vectors in $\mathbb{R}^2$.
3. What can you say about A^2 and other powers of A?
4. Is A invertible?

8

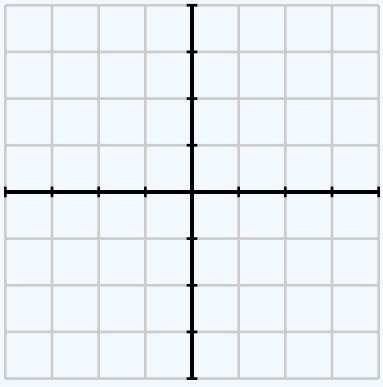
When A is a 2×2 matrix with a complex eigenvalue $\lambda = a + bi$, we have said that there is a matrix P such that $A = PCP^{-1}$ where $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. In this exercise, we will learn how to find the matrix P. As an example, we will consider the matrix $A = \begin{bmatrix} 2 & 2 \\ -1 & 4 \end{bmatrix}$.

1. Show that the eigenvalues of A are complex.
2. Choose one of the complex eigenvalues $\lambda = a + bi$ and construct the usual matrix C.
3. Using the same eigenvalue, we will find an eigenvector $\mathbf{v}$ where the entries of $\mathbf{v}$ are complex numbers. As always, we will describe $Nul(A - \lambda I)$ by constructing the matrix $A - \lambda I$ and finding its reduced row echelon form. In doing so, we will necessarily need to use complex arithmetic.
4. We have now found a complex eigenvector $\mathbf{v}$. Write $\mathbf{v} = \mathbf{v}_1 - i\mathbf{v}_2$ to identify vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ having real entries.
5. Construct the matrix $P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}$ and verify that $A = PCP^{-1}$.

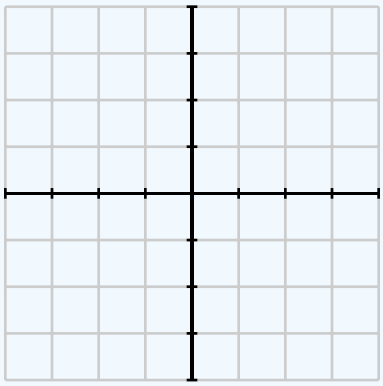
9

For each of the following matrices, sketch the vector $\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and powers $A^k\mathbf{x}$ for $k = 1, 2, 3, 4$.

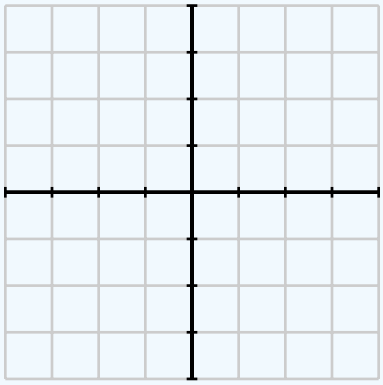
1. $A = \begin{bmatrix} 0 & -1.4 \\ 1.4 & 0 \end{bmatrix}$.



2. $A = \begin{bmatrix} 0 & -0.8 \\ 0.8 & 0 \end{bmatrix}$.



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

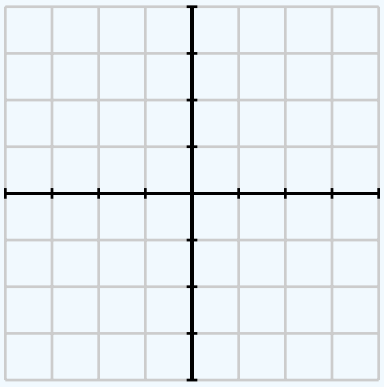


4. Consider a matrix of the form $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ with $r = \sqrt{a^2 + b^2}$. What happens when k becomes very large when
   1. $r < 1$.
   2. $r = 1$.
   3. $r > 1$.

10

For each of the following matrices and vectors, sketch the vector $\mathbf{x}$ along with $A^k\mathbf{x}$ for $k = 1, 2, 3, 4$.

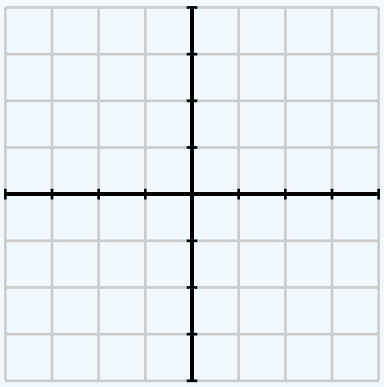
1. $$A = \begin{bmatrix} 1.4 & 0 \\ 0 & 0.7 \end{bmatrix}$$
   $$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$ (6.3.43)



2.

$$A = \begin{bmatrix} 0.6 & 0 \\ 0 & 0.9 \end{bmatrix} \qquad (6.3.44)$$

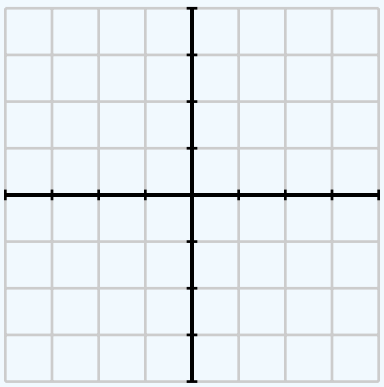
$$\mathbf{x} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}.$$



3.

$$A = \begin{bmatrix} 1.2 & 0 \\ 0 & 1.4 \end{bmatrix} \qquad (6.3.45)$$

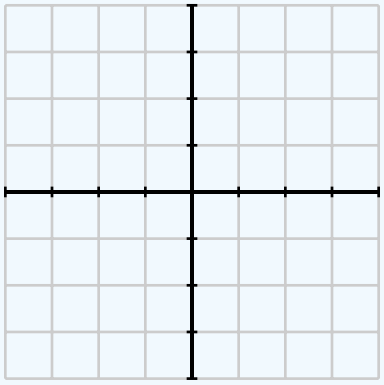
$$\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$



4.

$$A = \begin{bmatrix} 0.95 & 0.25 \\ 0.25 & 0.95 \end{bmatrix}$$

$$\mathbf{x} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}. \tag{6.3.46}$$



Find the eigenvalues and eigenvectors of A to create your sketch.

5. If A is a 2×2 matrix with eigenvalues $\lambda_1 = 0.7$ and $\lambda_2 = 0.5$ and $\mathbf{x}$ is any vector, what happens to $A^k\mathbf{x}$ when k becomes very large?

This page titled 6.3: Diagonalization, similarity, and powers of a matrix is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by David Austin via source content that was edited to the style and standards of the LibreTexts platform.

- 4.3: Diagonalization, similarity, and powers of a matrix by David Austin is licensed CC BY 4.0. Original source: https://davidaustinm.github.io/ula/ula.html.

6.4: Dynamical Systems

In the last section, we used a coordinate system defined by the eigenvectors of a matrix to express matrix multiplication in a simpler form. For instance, if there is a basis of $\mathbb{R}^n$ consisting of eigenvectors of A, we saw that multiplying a vector by A, when expressed in the coordinates defined by the basis of eigenvectors, was equivalent to multiplying by a diagonal matrix.

In this section, we will put these ideas to use as we explore discrete dynamical systems, first encountered in Subsection 2.5.2. Recall that we used a state vector $\mathbf{x}$ to characterize the state of some system, such as the distribution of delivery trucks between two locations, at a particular time. A matrix A described the transition of the state vector with $A\mathbf{x}$ characterizing the state of the system at a later time.

Our goal in this section is to describe the types of behaviors that dynamical systems exhibit and to develop a means of detecting these behaviors.

Preview Activity 4.4.1.

Suppose that we have a diagonalizable matrix $A = PDP^{-1}$ where

$$P = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}, \qquad D = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}.$$

1. Find the eigenvalues of A and find a basis for the associated eigenspaces.
2. Form a basis $\mathcal{B}$ of $\mathbb{R}^2$ consisting of eigenvectors of A and write the vector $\mathbf{x} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$ as a linear combination of basis vectors.
3. Write $A\mathbf{x}$ as a linear combination of basis vectors.
4. What is $\{\mathbf{x}\}_{\mathcal{B}}$, the representation of $\mathbf{x}$ in the coordinate system defined by $\mathcal{B}$?
5. What is $\{\mathbf{Ax}\}_{\mathcal{B}}$, the representation of $A\mathbf{x}$ in the coordinate system defined by $\mathcal{B}$?
6. What is $\{\mathbf{A}^4\mathbf{x}\}_{\mathcal{B}}$, the representation of $A^4\mathbf{x}$ in the coordinate system defined by $\mathcal{B}$?

A first example

We will begin our study of dynamical systems with an example that illustrates how eigenvalues and eigenvectors may be used to understand their behavior.

Activity 4.4.2.

Suppose we have two species R and S that interact with one another and that we record the change in their populations from year to year. When we begin our study, the populations, measured in thousands, are R_0 and S_0; after k years, the populations are R_k and S_k.

If we know the populations in one year, they are determined in the following year by the expressions

$$\begin{aligned} R_{k+1} &= 0.9R_k + 0.8S_k \\ S_{k+1} &= 0.2R_k + 0.9S_k. \end{aligned}$$

We will combine the populations in a vector $\mathbf{x}_k = \begin{pmatrix} R_k \\ S_k \end{pmatrix}$ and note that $\mathbf{x}_{k+1} = A\mathbf{x}_k$ where $A = \begin{bmatrix} 0.9 & 0.8 \\ 0.2 & 0.9 \end{bmatrix}$.

1. Verify that

$$\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

are eigenvectors of A and find their respective eigenvalues.

2. Suppose that initially $\mathbf{x}_0 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$. Write $\mathbf{x}_0$ as a linear combination of the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$.
3. Write the vectors $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ as a linear combination of eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$.
4. When k becomes very large, what happens to the ratio of the populations R_k/S_k?
5. If we begin instead with $\mathbf{x}_0 = \begin{pmatrix} 4 \\ 4 \end{pmatrix}$, what eventually happens to the ratio R_k/S_k as k becomes very large?
6. Explain what happens to the ratio R_k/S_k as k becomes very large no matter what the initial populations are.
7. After a very long time, by approximately what factor does the population of R grow every year? By approximately what factor does the population of S grow every year?

This activity demonstrates the type of questions we will be considering. In particular, we will assume that we have an initial vector $\mathbf{x}_0$ and a matrix A and define $\mathbf{x}_{k+1} = A\mathbf{x}_k$. The eigenvalues and eigenvectors of A provide the key that helps us understand how the vectors $\mathbf{x}_k$ evolve and enables us to make long-range predictions.

Let's look at the specific example in the previous activity more carefully. We see that we have

$$\mathbf{x}_{k+1} = A\mathbf{x}_k = \begin{bmatrix} 0.9 & 0.8 \\ 0.2 & 0.9 \end{bmatrix} \mathbf{x}_k$$

and that the matrix A has eigenvectors $\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$ with associated eigenvalues $\lambda_1 = 1.3$ and $\lambda_2 = 0.5$.

Notice that the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ form a basis $\mathcal{B}$ of $\mathbb{R}^2$. This means that A is diagonalizable so we can write $A = PDP^{-1}$ where

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 2 & -2 \\ 1 & 1 \end{bmatrix}, \qquad D = \begin{bmatrix} 1.3 & 0 \\ 0 & 0.5 \end{bmatrix}.$$

In particular, we have

$$\{A\mathbf{x}\}_{\mathcal{B}} = D\{\mathbf{x}\}_{\mathcal{B}}.$$

With intial populations $\mathbf{x}_0 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$, we have $\mathbf{x}_0 = 2\mathbf{v}_1 + \mathbf{v}_2$, which means that $\{\mathbf{x}_0\}_{\mathcal{B}} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Therefore,

$$\{\mathbf{x}_1\}_{\mathcal{B}} = D\{\mathbf{x}_0\}_{\mathcal{B}} = \begin{bmatrix} 1.3 \cdot 2 \\ 0.5 \cdot 1 \end{bmatrix}$$
$$\{\mathbf{x}_2\}_{\mathcal{B}} = D\{\mathbf{x}_1\}_{\mathcal{B}} = \begin{bmatrix} 1.3^2 \cdot 2 \\ 0.5^2 \cdot 1 \end{bmatrix}$$
$$\{\mathbf{x}_3\}_{\mathcal{B}} = D\{\mathbf{x}_2\}_{\mathcal{B}} = \begin{bmatrix} 1.3^3 \cdot 2 \\ 0.5^3 \cdot 1 \end{bmatrix}$$
$$\{\mathbf{x}_k\}_{\mathcal{B}} = \begin{bmatrix} 1.3^k \cdot 2 \\ 0.5^k \cdot 1 \end{bmatrix}.$$

Thinking about this geometrically, we begin with the vector $\{\mathbf{x}_0\}_{\mathcal{B}} = \begin{bmatrix} 1.3^2 \cdot 2 \\ 0.5^k \cdot 1 \end{bmatrix}$. Subsequent vectors $\{\mathbf{x}_k\}_{\mathcal{B}}$ are obtained by scaling horizontally by a factor of 1.3 and scaling vertically by a factor 0.5. Notice how the points move along a curve away from the origin becoming ever closer to the horizontal axis.



Figure 6.4.1: Paste Caption Here

After a very long time, $\{\mathbf{x}_k\}_{\mathcal{B}} \approx \begin{bmatrix} 1.3^k \cdot 2 \\ 0 \end{bmatrix}$, which says that $\{\mathbf{x}_{k+1}\}_{\mathcal{B}} \approx 1.3\{\mathbf{x}_k\}_{\mathcal{B}}$. That is, the vector $(\{\mathbf{x}_k\}_{\mathcal{B}}$ grows by a factor of 1.3 every year.

To recover the behavior of the sequence $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$, we change coordinate systems using the basis defined by $\mathbf{v}_1$ and $\mathbf{v}_2$. Here, the points move along a curve away from the origin becoming ever closer to the line defined by $\mathbf{v}_1$.



Figure 6.4.1: Paste Caption Here

Eventually, the vectors become practically indistinguishable from a scalar multiple of $\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$; that is, $\mathbf{x}_k \approx s\mathbf{v}_1$. This means that

$$\mathbf{x}_k = \begin{pmatrix} R_k \\ S_k \end{pmatrix} \approx s \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

so that $R_k/S_k \approx 2$. In addition, $\mathbf{x}_{k+1} \approx 1.3\mathbf{x}_k$ so that $R_{k+1} \approx 1.3R_k$ and $S_{k+1} \approx 1.3S_k$. We conclude that, after a very long time, the ratio of the populations R_k to S_k is very close to 2 to 1. We also see that each population is multiplied by 1.3 every year meaning the annual growth rate for both populations is about 30%.

In the same way, we can consider other possible initial populations $\mathbf{x}_0$ as shown in Figure 4.4.1. Regardless of $\mathbf{x}_0$, the population vectors, in the coordinates defined by $\mathcal{B}$, are scaled horizontally by a factor of 1.3 and vertically by a factor of 0.5. The sequence of points $\{\mathbf{x}_k\}_{\mathcal{B}}$, called *trajectories*, move along the curves, as shown on the left. In the standard coordinate system, we see that the trajectories converge to the eigenspace $E_{1.3}$.

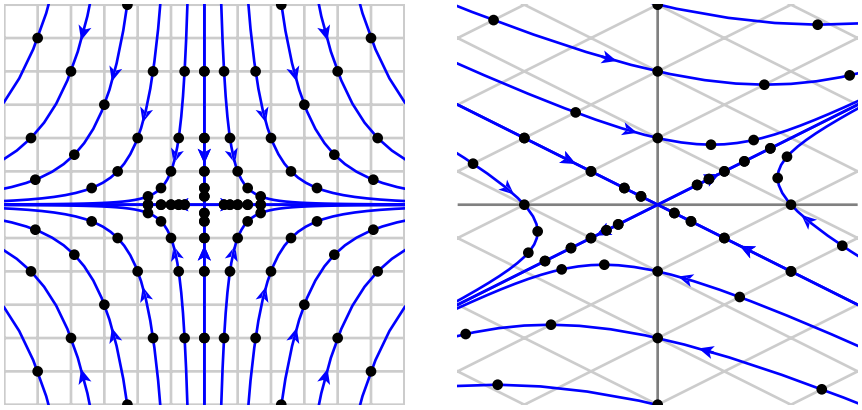


Figure 6.4.1: The trajectories of the dynamical system formed by the matrix A in the coordinate system defined by $\mathcal{B}$, on the left, and in the standard coordinate system, on the right.

We conclude that, regardless of the initial populations, the ratio of the populations R_k/S_k will approach 2 to 1 and that the growth rate for both populations approaches 30%. This example demonstrates the power of using eigenvalues and eigenvectors to rewrite the problem in terms of a new coordinate system. By doing so, we are able to predict the long-term behavior of the populations independently of the initial populations.

Diagrams like those shown in Figure 4.4.1 are called phase portraits. On the left of Figure 4.4.1 is the phase portrait of the diagonal matrix $D = \begin{bmatrix} 1.3 & 0 \\ 0 & 0.5 \end{bmatrix}$ while the right of that figure shows the phase portrait of $A = \begin{bmatrix} 0.9 & 0.8 \\ 0.2 & 0.9 \end{bmatrix}$. The phase portrait of D is relatively easy to understand because it is determined only by the two eigenvalues. Once we have the phase portrait of D, however, the phase portrait of A has a similar appearance with the eigenvectors $\mathbf{v}_j$ replacing the standard basis vectors $\mathbf{e}_j$.

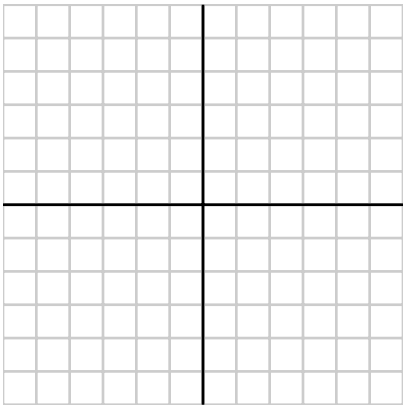
Classifying dynamical systems

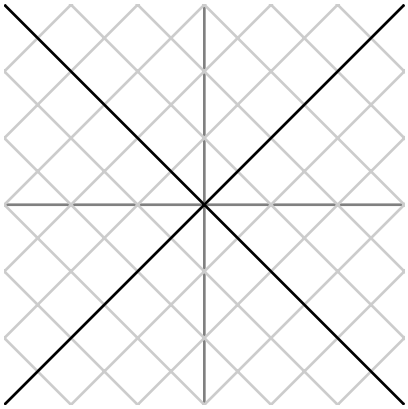
In the previous example, we were able to make predictions about the behavior of trajectories $\mathbf{x}_k = A^k\mathbf{x}_0$ by considering the eigenvalues and eigenvectors of the matrix A. The next activity looks at a collection of matrices that demonstrate the types of behavior a 2×2 dynamical system can exhibit.

Activity 4.4.3.

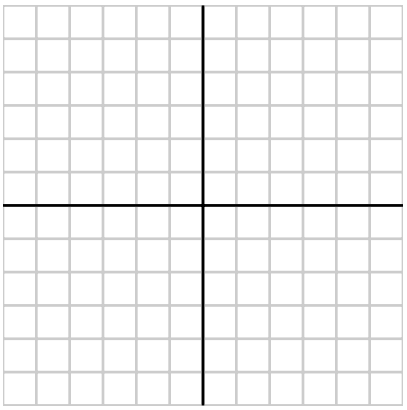
We will now look at several more examples of dynamical systems. If $P = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, we note that the columns of P form a basis $\mathcal{B}$ of $\mathbb{R}^2$. Given below are several matrices A written in the form $A = PEP^{-1}$ for some matrix E. For each matrix, state the eigenvalues of A and sketch a phase portrait for the matrix E on the left and a phase portrait for A on the right. Describe the behavior of $A^k\mathbf{x}_0$ as k becomes very large for a typical initial vector $\mathbf{x}_0$.

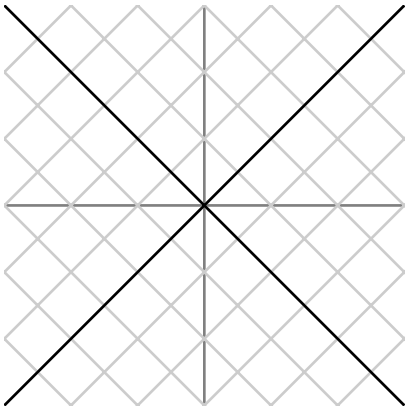
1. $A = PEP^{-1}$ where $E = \begin{bmatrix} 1.3 & 0 \\ 0 & 1.5 \end{bmatrix}$.



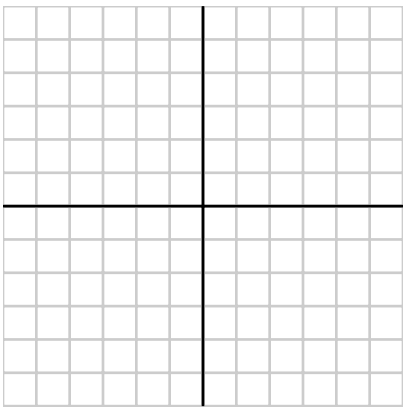


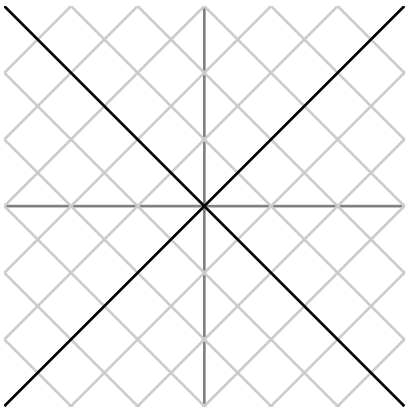
2. $A = PEP^{-1}$ where $E = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.



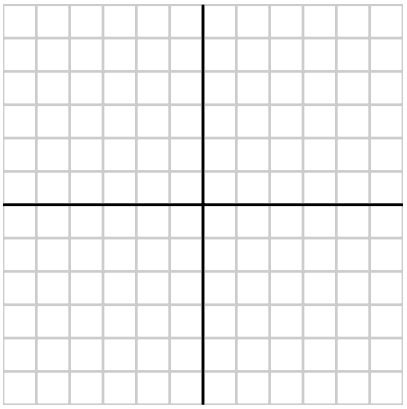


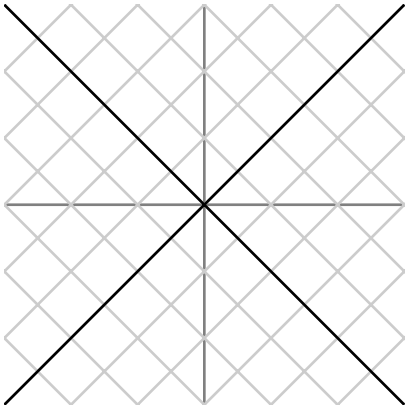
3. $A = PEP^{-1}$ where $E = \begin{bmatrix} 0.7 & 0 \\ 0 & 1.5 \end{bmatrix}$.



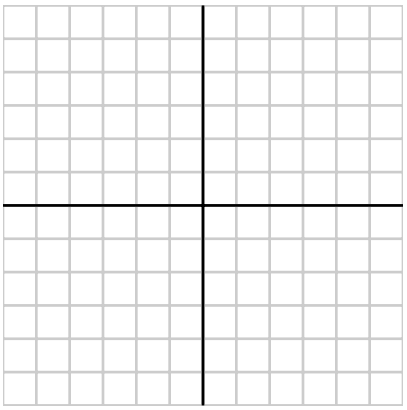


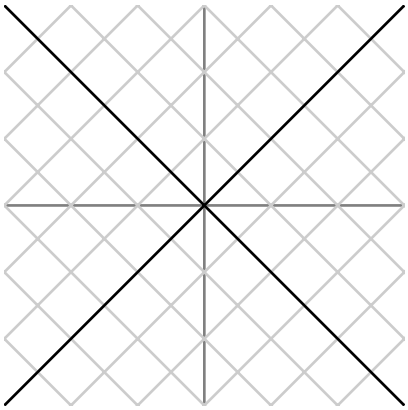
4. $A = PEP^{-1}$ where $E = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.7 \end{bmatrix}$.



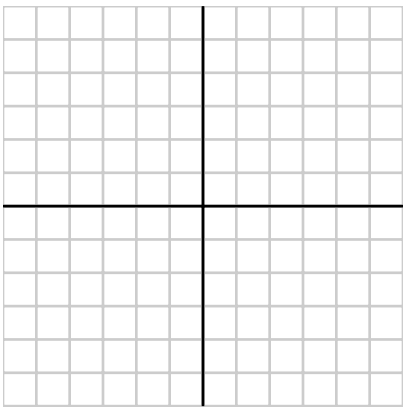


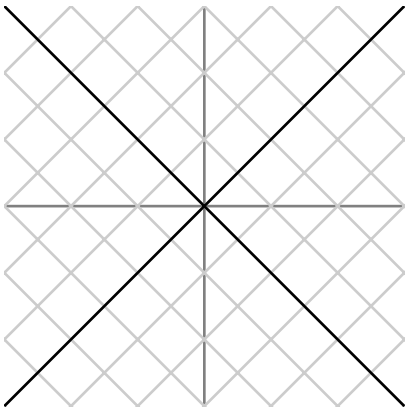
5. $A = PEP^{-1}$ where $E = \begin{bmatrix} 1 & -0.9 \\ 0.9 & 1 \end{bmatrix}$.





6. $A = PEP^{-1}$ where $E = \begin{bmatrix} 0.6 & -0.2 \\ 0.2 & 0.6 \end{bmatrix}$.





This activity demonstrates six possible types of dynamical systems, which are determined by the eigenvalues of A.

- Suppose that A has two real eigenvalues λ_1 and λ_2 and that both $|\lambda_1|, |\lambda_2| > 1$. In this case, any nonzero vector $\mathbf{x}_0$ forms a trajectory that moves away from the origin so we say that the origin is a *repellor*. This is illustrated in Figure 4.4.2.

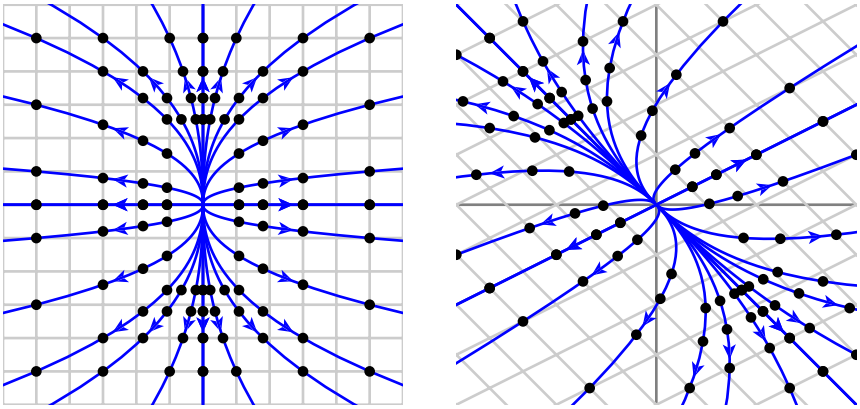


Figure 6.4.2: The origin is a repellor when $|\lambda_1|, |\lambda_2| > 1$.

- Suppose that A has two real eigenvalues λ_1 and λ_2 and that $|\lambda_1| > 1 > |\lambda_2|$. In this case, most nonzero vectors $\mathbf{x}_0$ form trajectories that converge to the eigenspace E_{λ_1}. In this case, we say that the origin is a *saddle* as illustrated in Figure 4.4.3.

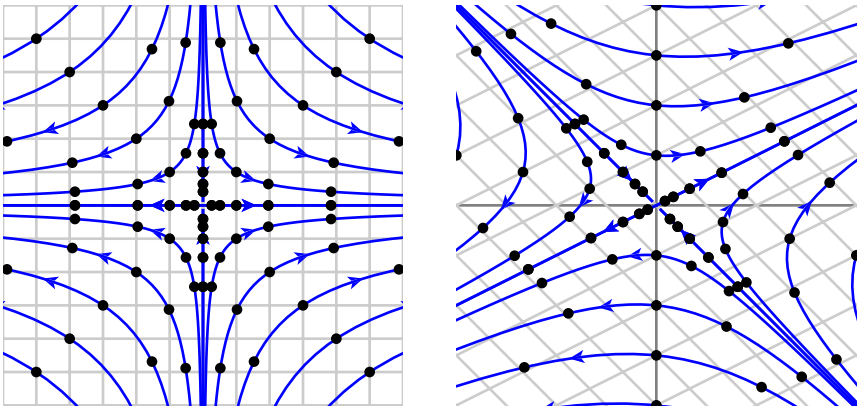


Figure 6.4.3: The origin is a saddle when $|\lambda_1| > 1 > |\lambda_2|$.

- Suppose that A has two real eigenvalues λ_1 and λ_2 and that both $|\lambda_1|, |\lambda_2| < 1$. In this case, any nonzero vector $\mathbf{x}_0$ forms a trajectory that moves into the origin so we say that the origin is an *attractor*. This is illustrated in Figure 4.4.4.

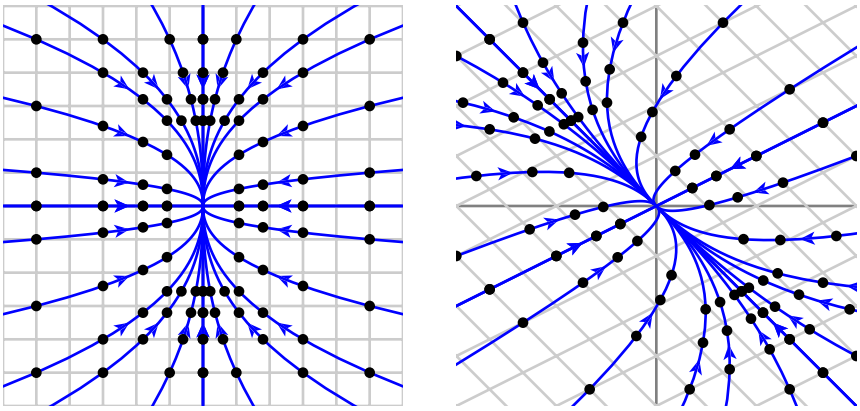


Figure 6.4.4: The origin is an attractor when $|\lambda_1|, |\lambda_2| < 1$.

- Suppose that A has a complex eigenvalue $\lambda = a + bi$ where $|\lambda| > 1$. In this case, a nonzero vector $\mathbf{x}_0$ forms a trajectory that spirals away from the origin. We say that the origin is a *spiral repellor*, as illustrated in Figure 4.4.5.

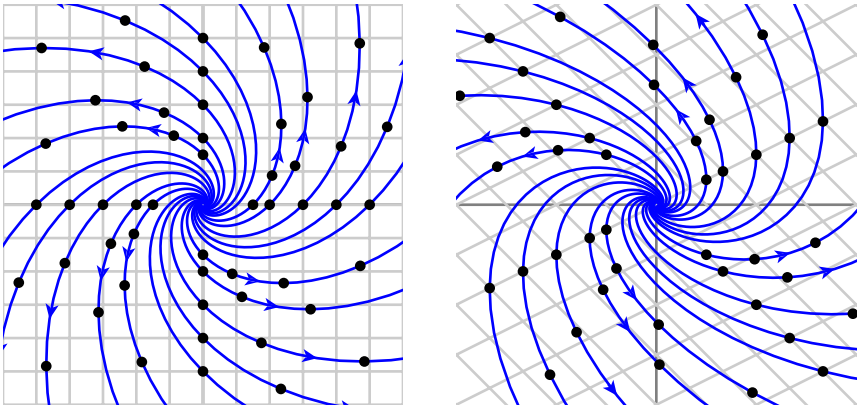


Figure 6.4.5: The origin is a spiral repellor when A has an eigenvalue $\lambda = a + bi$ with $a^2 + b^2 > 1$.

- Suppose that A has a complex eigenvalue $\lambda = a + bi$ where $|\lambda| = 1$. In this case, a nonzero vector $\mathbf{x}_0$ forms a trajectory that moves on a closed curve around the origin. We say that the origin is a *center*, as illustrated in Figure 4.4.6.

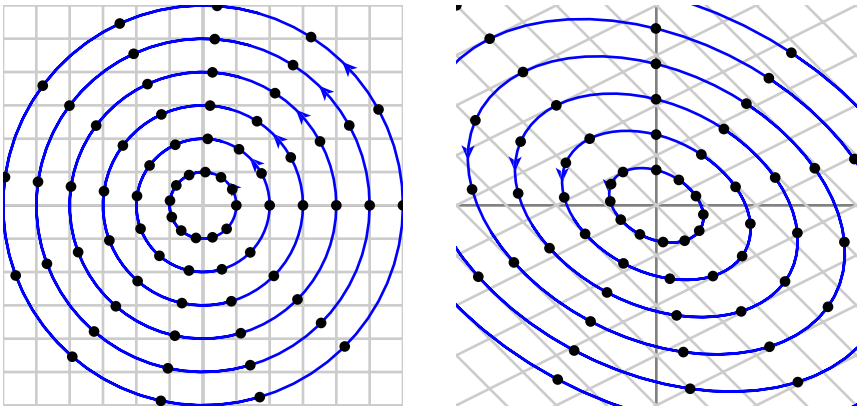


Figure 6.4.6: The origin is a center when A has an eigenvalue $\lambda = a + bi$ with $a^2 + b^2 = 1$.

- Suppose that A has a complex eigenvalue $\lambda = a + bi$ where $|\lambda| < 1$. In this case, a nonzero vector $\mathbf{x}_0$ forms a trajectory that spirals into the origin. We say that the origin is a *spiral attractor*, as illustrated in Figure 4.4.7.

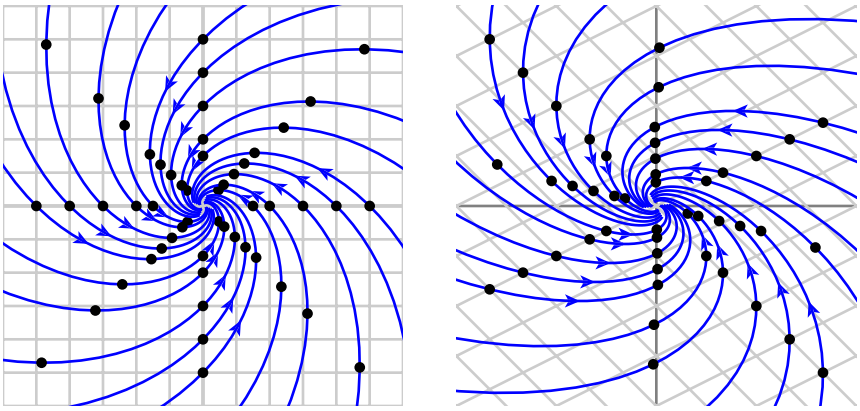


Figure 6.4.7: The origin is a spiral attractor when A has an eigenvalue $\lambda = a + bi$ with $a^2 + b^2 < 1$.

Activity 4.4.4.

In this activity, we will consider several ways in which two species might interact with one another. Throughout, we will consider two species R and S whose populations in year k form a vector $\mathbf{x}_k = \begin{pmatrix} R_k \\ S_k \end{pmatrix}$ and which evolve according to the rule

$$\mathbf{x}_{k+1} = A\mathbf{x}_k.$$

1. Suppose that $A = \begin{bmatrix} 0.7 & 0 \\ 0 & 1.6 \end{bmatrix}$.

   Explain why the species do not interact with one another. Which of the six types of dynamical systems do we have? What happens to both species after a long time?

2. Suppose now that $A = \begin{bmatrix} 0.7 & 0.3 \\ 0 & 1.6 \end{bmatrix}$.

   Explain why S is a beneficial species for R. Which of the six types of dynamical systems do we have? What happens to both species after a long time?

3. Suppose now that $A = \begin{bmatrix} 0.7 & 0.5 \\ -0.4 & 1.6 \end{bmatrix}$.

   Explain why this describes a predator-prey system. Which of the species is the predator and which is the prey? Which of the six types of dynamical systems do we have? What happens to both species after a long time?

4. Suppose now that $A = \begin{bmatrix} 0.5 & 0.2 \\ -0.4 & 1.1 \end{bmatrix}$.

   Compare this predator-prey system to the one in the previous part. Which of the six types of dynamical systems do we have? What happens to both species after a long time?

A 3×3 system

Up to this point, we have focused on 2×2 systems. In fact, the general case is quite similar. As an example, consider a 3×3 system $\mathbf{x}_{k+1} = A\mathbf{x}_k$ where the matrix A has eigenvalues $\lambda_1 = 0.6$, $\lambda_2 = 0.8$, and $\lambda_3 = 1.1$. The matrix A therefore has a basis $\mathcal{B}$ consisting of eigenvectors so we can look at the trajectories $\{\mathbf{x}_k\}_{\mathcal{B}}$ in the coordinate system defined by $\mathcal{B}$. The phase portraits in Figure 4.4.8 show how the trajectories will evolve. We see that all the trajectories will converge into the eigenspace $E_{1.1}$.

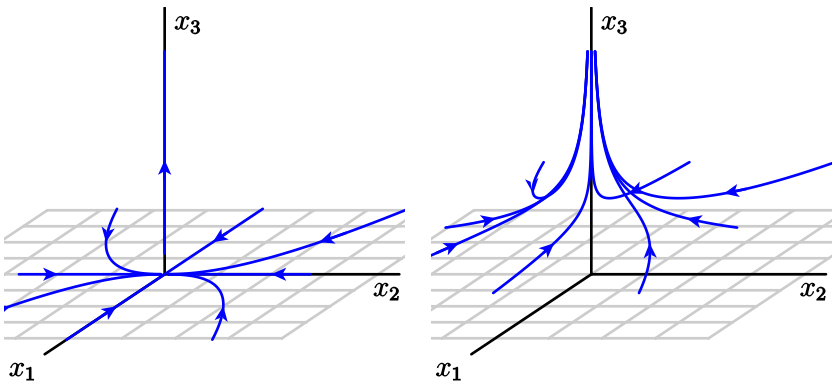


Figure 6.4.8: In a 3×3 system with $\lambda_1 = 0.6$, $\lambda_2 = 0.8$, and $\lambda_3 = 1.1$, the trajectories $\{\mathbf{x}_k\}_{\mathcal{B}}$ move along the curves shown above.

In the same way, suppose we have a 3×3 system with complex eigenvalues $\lambda = 0.8 \pm 0.5i$ and $\lambda_3 = 1.1$. Since the complex eigenvalues satisfy $|\lambda| < 1$, there is a two-dimensional subspace in which the trajectories spiral in toward the origin. The phase portraits in Figure 4.4.9 show some of the trajectories. Once again, we see that all the trajectories converge into the eigenspace $E_{1.1}$.

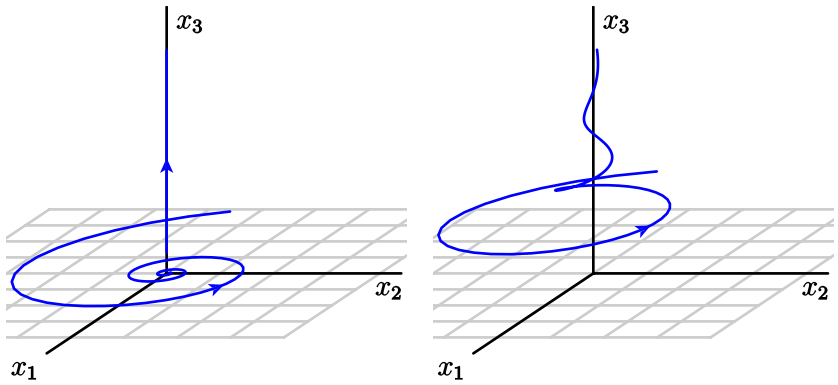


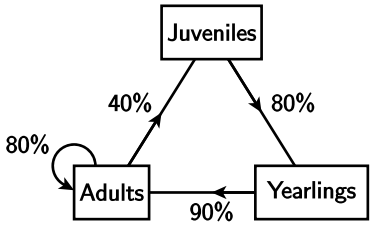
Figure 6.4.9: In a 3×3 system with complex eigenvalues $\lambda = a \pm bi$ with $|\lambda| < 1$ and $\lambda_3 = 1.1$, the trajectories $\{\mathbf{x}_k\}_{\mathcal{B}}$ move along the curves shown above.

Activity 4.4.5.

The following type of analysis has been used to study the population of a bison herd. We will divide the population of female bison into three groups: juveniles who are less than one year old; yearlings between one and two years old; and adults who are older than two years.

Each year,

- 80% of the juveniles survive to become yearlings.
- 90% of the yearlings survive to become adults.
- 80% of the adults survive.
- 40% of the adults give birth to a juvenile.



By J_k, Y_k, and A_k, we denote the number of juveniles, yearlings, and adults in year k. We have

$$J_{k+1} = 0.4A_k.$$

1. Find similar expressions for Y_{k+1} and A_{k+1} in terms of J_k, Y_k, and A_k.
2. As is usual, we write the matrix $\mathbf{x}_k = \begin{pmatrix} J_k \\ Y_k \\ A_k \end{pmatrix}$. Write the matrix A such that $\mathbf{x}_{k+1} = A\mathbf{x}_k$.
3. We can write $A = PEP^{-1}$ where the matrices E and P are approximately:

$$E = \begin{bmatrix} 1.058 & 0 & 0 \\ 0 & -0.128 & -0.506 \\ 0 & 0.506 & -0.128 \end{bmatrix},$$

$$P = \begin{bmatrix} 1 & 1 & 0 \\ 0.756 & -0.378 & 1.486 \\ 2.644 & -0.322 & -1.264 \end{bmatrix}.$$

Make a prediction about the long-term behavior of $\mathbf{x}_k$. For instance, at what rate does it grow? For every 100 adults, how many juveniles, and yearlings are there?

4. Suppose that the birth rate decreases so that only 30% of adults give birth to a juvenile. How does this affect the long-term growth rate of the herd?
5. Suppose that the birth rate decreases further so that only 20% of adults give birth to a juvenile. How does this affect the long-term growth rate of the herd?
6. Find the smallest birth rate that supports a stable population.

Summary

We have been exploring discrete dynamical systems, which have the form $\mathbf{x}_{k+1} = A\mathbf{x}_k$, by looking at the eigenvalues and eigenvectors of A. In the 2×2 case, we saw that

- $|\lambda_1|, |\lambda_2| < 1$ produces an attractor so that trajectories are pulled in toward the origin.
- $|\lambda_1| > 1$ and $|\lambda_2| < 1$ produces a saddle in which most trajectories are pushed away from the origin and in the direction of E_{λ_1}.
- $|\lambda_1|, |\lambda_2| > 1$ produces a repellor in which trajectories are pushed away from the origin.

The same kind of reasoning allows us to analyze $n \times n$ systems as well.

Exercises 4.4.5Exercises

1

For each of the 2×2 matrices below, find the eigenvalues and, when appropriate, the eigenvectors to classify the dynamical system $\mathbf{x}_{k+1} = A\mathbf{x}_k$. Use this information to sketch the phase portraits.

1. $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$.
2. $A = \begin{bmatrix} 3 & -2 \\ 4 & -1 \end{bmatrix}$.
3. $A = \begin{bmatrix} 1.9 & 1.4 \\ -0.7 & -0.2 \end{bmatrix}$.
4. $A = \begin{bmatrix} 1.1 & -0.2 \\ 0.4 & 0.5 \end{bmatrix}$.

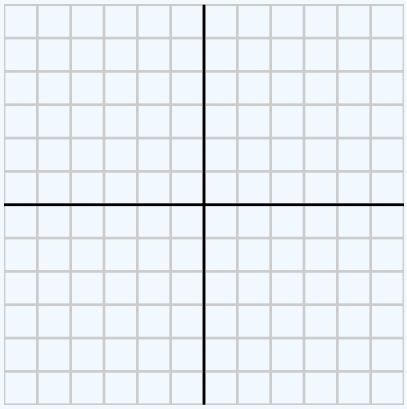
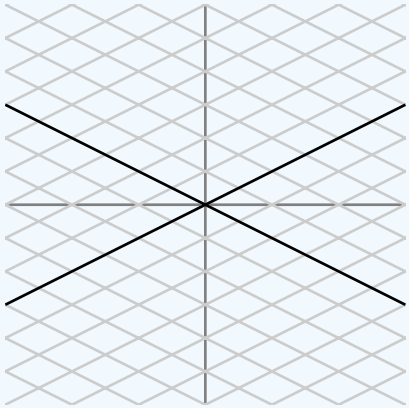
2

We will consider matrices that have the form $A = PDP^{-1}$ where

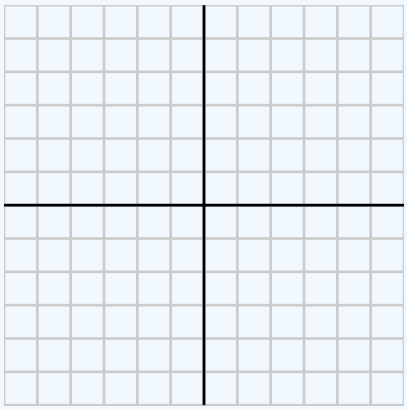
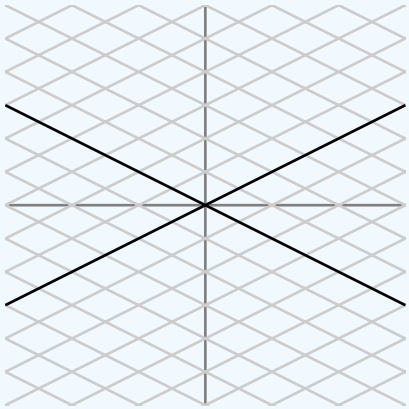
$$D = \begin{bmatrix} p & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, P = \begin{bmatrix} 2 & -2 \\ 1 & 1 \end{bmatrix}$$

where p is a parameter that we will vary. Sketch phase portraits for D and A below when

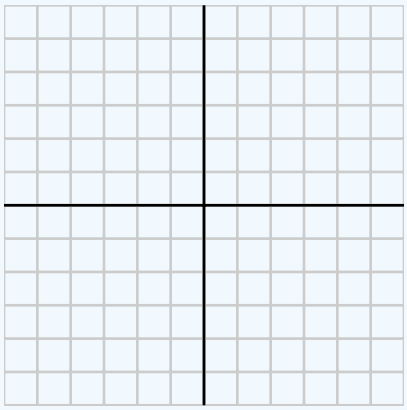
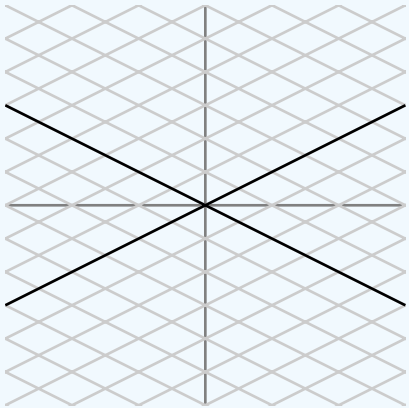
1. $p = \frac{1}{2}$.

2. $p = 1$.

3. $p = 2$.

4. For the different values of p, determine which types of dynamical system results. For what range of p values do we have an attractor? For what range of p values do we have a saddle? For what value does the transition between the two types occur?

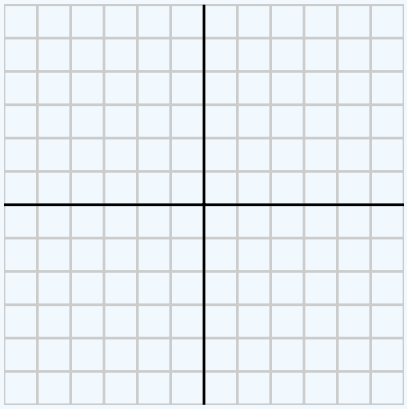
3

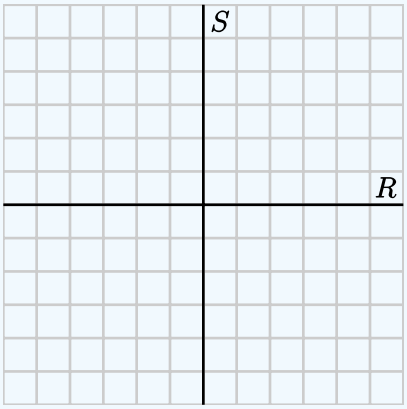
Suppose that the populations of two species interact according to the relationships

$$\begin{aligned} R_{k+1} &= \frac{1}{2}R_k + \frac{1}{2}S_k \\ S_{k+1} &= -pR_k + 2S_k \end{aligned}$$

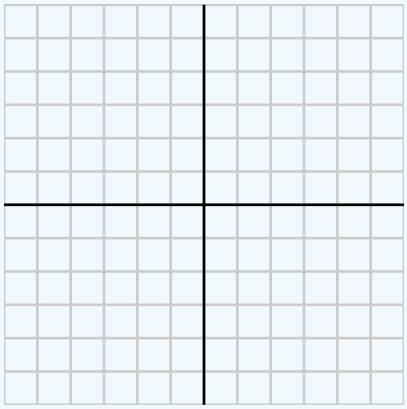
where p is a parameter. As we saw in the text, this dynamical system represents a typical predator-prey relationship, and the parameter p represents the rate at which species R preys on S. We will denote the matrix $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -p & 2 \end{bmatrix}$.

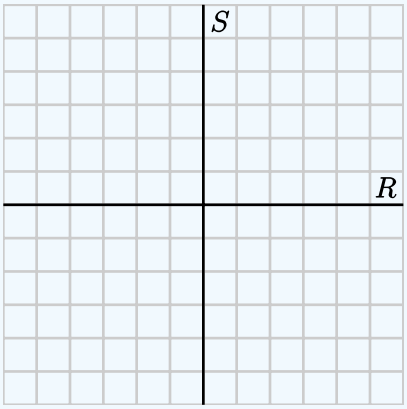
1. If $p = 0$, determine the eigenvectors and eigenvalues of the system and classify it as one of the six types. Sketch the phase portraits for the diagonal matrix D to which A is similar as well as the phase portrait for A.





2. If $p = 1$, determine the eigenvectors and eigenvalues of the system. Sketch the phase portraits for the diagonal matrix D to which A is similar as well as the phase portrait for A.





3. For what values of p is the origin a saddle? What can you say about the populations when this happens?
4. Describe the evolution of the dynamical system as p begins at 0 and increases to $p = 1$.

4

Consider the matrices

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

1. Find the eigenvalues of A. To which of the six types does the system $\mathbf{x}_{k+1} = A\mathbf{x}_k$ belong?
2. Using the eigenvalues of A, we can write $A = PEP^{-1}$ for some matrices E and P. What is the matrix E and what geometric effect does multiplication by E have on vectors in the plane?
3. If we remember that $A^k = PE^kP^{-1}$, determine the smallest positive value of k for which $A^k = I$?
4. Find the eigenvalues of B.
5. Then find a matrix E such that $B = PEP^{-1}$ for some matrix P. What geometric effect does multiplication by E have on vectors in the plane?
6. Determine the smallest positive value of k for which $B^k = I$.

5

Suppose we have the female population of a species is divided into juveniles, yearlings, and adults and that each year

- 90% of the juveniles live to be yearlings.
- 80% of the yearlings live to be adults.
- 60% of the adults survive to the next year.
- 50% of the adults give birth to a juvenile.

1. Set up a system of the form $\mathbf{x}_{k+1} = A\mathbf{x}_k$ that describes this situation.
2. Find the eigenvalues of the matrix A.
3. What prediction can you make about these populations after a very long time?
4. If the birth rate goes up to 80%, what prediction can you make about these populations after a very long time? For every 100 adults, how many juveniles, and yearlings are there?

6

Determine whether the following statements are true or false and provide a justification for your response. In each case, we are considering a dynamical system of the form $\mathbf{x}_{k+1} = A\mathbf{x}_k$.

1. If the 2×2 matrix A has a complex eigenvalue, we cannot make a prediction about the behavior of the trajectories.
2. If A has eigenvalues whose absolute value is smaller than 1, then all the trajectories are pulled in toward the origin.
3. If the origin is a repellor, then it is an attractor for the system $\mathbf{x}_{k+1} = A^{-1}\mathbf{x}_k$.
4. If a 4×4 matrix has complex eigenvalues λ_1, λ_2, λ_3, and λ_4, all of which satisfy $|\lambda_j| > 1$, then all the trajectories are pushed away from the origin.
5. If the origin is a saddle, then all the trajectories are pushed away from the origin.

7

The Fibonacci numbers form the sequence of numbers that begins $0, 1, 1, 2, 3, 5, 8, 13, \ldots$. If we let F_n denote the nth Fibonacci number, then

$$F_0 = 0, F_1 = 1, F_2 = 1, F_3 = 2, F_4 = 3, \ldots.$$

In general, a Fibonacci number is the sum of the previous two Fibonacci numbers; that is, $F_{n+2} = F_n + F_{n+1}$ so that we have

$$\begin{aligned} F_{n+2} &= F_n + F_{n+1} \\ F_{n+1} &= F_{n+1}. \end{aligned}$$

1. If we write $\mathbf{x}_n = \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix}$, find the matrix A such that $\mathbf{x}_{n+1} = A\mathbf{x}_n$.
2. Show that A has eigenvalues

$$\lambda_1 = \frac{1+\sqrt{5}}{2} \approx 1.61803\ldots$$
$$\lambda_2 = \frac{1-\sqrt{5}}{2} \approx -0.61803\ldots$$

with associated eigenvectors $\mathbf{v}_1 = \begin{pmatrix}\lambda_1\\1\end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix}\lambda_2\\1\end{pmatrix}$.

3. Classify this dynamical system as one of the six types that we have seen in this section. What happens to $\mathbf{x}_n$ as n becomes very large?
4. Write the initial vector $\mathbf{x}_0 = \begin{pmatrix}1\\0\end{pmatrix}$ as a linear combination of eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$.
5. Write the vector $\mathbf{x}_n$ as a linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$.
6. Explain why the nth Fibonacci number
$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right].$$
7. Use this relationship to compute F_{20}.
8. Explain why $F_{n+1}/F_n \approx \lambda_1$ when n is very large.

The number $\lambda_1 = \frac{1+\sqrt{5}}{2} = \phi$ is called the *golden ratio* and is one of mathematics' special numbers.

8

This exercise is a continuation of the previous one.

The Lucas numbers L_n are defined by the same relationship as the Fibonacci numbers: $L_{n+2} = L_{n+1} + L_n$. However, we begin with $L_0 = 2$ and $L_1 = 1$, which leads to the sequence $2, 1, 3, 4, 7, 11, \ldots$.

1. As before, form the vector $\mathbf{x}_n = \begin{pmatrix}L_{n+1}\\L_n\end{pmatrix}$ so that $\mathbf{x}_{n+1} = A\mathbf{x}_n$. Express $\mathbf{x}_0$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$, eigenvectors of A.
2. Explain why
$$L_n = \left(\frac{1+\sqrt{5}}{2}\right)^n + \left(\frac{1-\sqrt{5}}{2}\right)^n.$$
3. Explain why L_n is the closest integer to ϕ^n when n is large, where $\phi = \lambda_1$ is the golden ratio.
4. Use this observation to find L_{20}.

9

Gil Strang defines the *Gibonacci numbers* G_n as follows. We begin with $G_0 = 0$ and $G_1 = 1$. A subsequent Gibonacci number is the average of the two previous; that is, $G_{n+2} = \frac{1}{2}(G_n + G_{n+1})$. We then have

$$G_{n+2} = \frac{1}{2}G_n + \frac{1}{2}G_{n+1}$$
$$G_{n+1} = G_{n+1}.$$

1. If $\mathbf{x}_n = \begin{pmatrix}G_{n+1}\\G_n\end{pmatrix}$, find the matrix A such that $\mathbf{x}_{n+1} = A\mathbf{x}_n$.
2. Find the eigenvalues and associated eigenvectors of A.
3. Explain why this dynamical system does not neatly fit into one of the six types that we saw in this section.
4. Write $\mathbf{x}_0$ as a linear combination of eigenvectors of A.
5. Write $\mathbf{x}_n$ as a linear combination of eigenvectors of A.
6. What happens to G_n as n becomes very large?

10

Consider a small rodent that lives for three years. Once again, we can separate a population of females into juveniles, yearlings, and adults. Suppose that, each year,

- Half of the juveniles live to be yearlings.
- One quarter of the yearlings live to be adults.
- Adult females produce eight female offspring.
- None of the adults survive to the next year.

1. Writing the populations of juveniles, yearlings, and adults in year k using the vector $\mathbf{x}_k = \begin{pmatrix} J_k \\ Y_k \\ A_k \end{pmatrix}$, find the matrix A such that $\mathbf{x}_{k+1} = A\mathbf{x}_k$.
2. Show that $A^3 = I$.
3. What are the eigenvalues of A^3? What does this say about the eigenvalues of A?
4. Verify your observation by finding the eigenvalues of A.
5. What can you say about the trajectories of this dynamical system?
6. What does this mean about the population of rodents?
7. Find a population vector $\mathbf{x}_0$ that is unchanged from year to year.

This page titled 6.4: Dynamical Systems is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by David Austin via source content that was edited to the style and standards of the LibreTexts platform.

- 4.4: Dynamical Systems by David Austin is licensed CC BY 4.0. Original source: https://davidaustinm.github.io/ula/ula.html.

6.5: Markov chains and Google's PageRank algorithm

In the last section, we used our understanding of eigenvalues and eigenvectors to describe the long-term behavior of some discrete dynamical systems. The state of the sytem, which could record, say, the populations of a few interacting species, at one time was described by a vector $\mathbf{x}_k$. The state vector then evolved according to a linear rule $\mathbf{x}_{k+1} = A\mathbf{x}_k$.

We continue our exploration in this section by looking at Markov chains, which form a specific type of discrete dynamical system. For instance, we could be interested in a rental car company that rents cars from several locations. From one day to the next, the number of cars at different locations can change, but the total number of cars stays the same. Once again, an understanding of eigenvalues and eigenvectors will help us make predictions about the long-term behavior of the system.

Preview Activity 4.5.1.

Suppose that our rental car company rents from two locations P and Q. We find that 80% of the cars rented from location P are returned to P while the other 20% are returned to Q. For cars rented from location Q, 60% are returned to Q and 40% to P.

We will use P_k and Q_k to denote the number of cars at the two locations on day k. The following day, the number of cars at P equals 80% of P_k and 40% of Q_k. This shows that

$$\begin{aligned} P_{k+1} &= 0.8P_k + 0.4Q_k \\ Q_{k+1} &= 0.2P_k + 0.6Q_k. \end{aligned}$$

1. If we use the vector $\mathbf{x}_k = \begin{pmatrix} P_k \\ Q_k \end{pmatrix}$ to represent the distribution of cars on day k, find a matrix A such that $\mathbf{x}_{k+1} = A\mathbf{x}_k$.
2. Find the eigenvalues and associated eigenvectors of A.
3. Suppose that there are initially 1500 cars, all of which are at location P. Write the vector $\mathbf{x}_0$ as a linear combination of eigenvectors of A.
4. Write the vectors $\mathbf{x}_k$ as a linear combination of eigenvectors of A.
5. What happens to the distribution of cars after a long time?

A first example

In the preview activity, the distribution of rental cars was described by the discrete dynamical system

$$\mathbf{x}_{k+1} = A\mathbf{x}_k = \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{bmatrix} \mathbf{x}_k.$$

This matrix has some special properties. First, each entry represents the probability that a car rented at one location is returned to another. For instance, there is an 80% chance that a car rented at P is returned to P, which explains the entry of 0.8 in the upper left corner. Therefore, the entries of the matrix are between 0 and 1.

Second, a car rented at one location must be returned to one of the locations. For example, since 80% of the cars rented at P are returned to P, it follows that the other 20% of cars rented at P are returned to Q. This implies that the entries in each column must add to 1. This will occur frequently in our discussion so we introduce the following definitions.

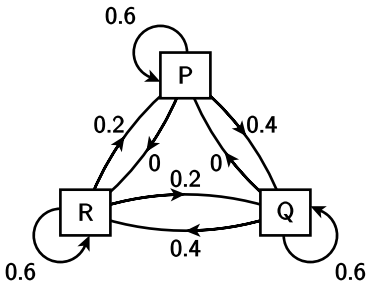
Definition 4.5.1

A vector whose entries are nonnegative and add to 1 is called a *probability vector*. A square matrix whose columns are probability vectors is called a *stochastic* matrix.

Activity 4.5.2.

Suppose you live in a country with three political parties P, Q, and R. We use P_k, Q_k, and R_k to denote the percentage of voters voting for that party in election k.

Voters will change parties from one election to the next as shown in the figure. We see that 60% of voters stay with the same party. However, 40% of those who vote for party P will vote for party Q in the next election.



1. Write expressions for P_{k+1}, Q_{k+1}, and R_{k+1} in terms of P_k, Q_k, and R_k.
2. If we write $\mathbf{x}_k = \begin{pmatrix} P_k \\ Q_k \\ R_k \end{pmatrix}$, find the matrix A such that $\mathbf{x}_{k+1} = A\mathbf{x}_k$.
3. Explain why A is a stochastic matrix.
4. Suppose that initially 40% of citizens vote for party P, 30% vote for party Q, and 30% vote for party R. Form the vector $\mathbf{x}_0$ and explain why $\mathbf{x}_0$ is a probability vector.
5. Find $\mathbf{x}_1$, the percentages who vote for the three parties in the next election. Verify that $\mathbf{x}_1$ is also a probabilty vector and explain why $\mathbf{x}_k$ will be a probability vector for every k.
6. Find the eigenvalues of the matrix A and explain why the eigenspace E_1 is a one-dimensional subspace of $\mathbb{R}^3$. Then verify that $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$ is a basis vector for E_1.
7. As every vector in E_1 is a scalar multiple of $\mathbf{v}$, find a probability vector in E_1 and explain why it is the only probability vector in E_1.
8. Describe what happens to $\mathbf{x}_k$ after a very long time.

The previous activity illustrates some important points that we wish to emphasize.

First, to determine P_{k+1}, we note that in election $k+1$, party P retains 60% of its voters from the previous election and adds 20% of those who voted for party R. In this way, we see that

$$\begin{aligned} P_{k+1} &= 0.6P_k &&+0.2R_k \\ Q_{k+1} &= 0.4P_k + 0.6Q_k + 0.2R_k \\ R_{k+1} &= 0.4Q_k + 0.6R_k \end{aligned}$$

We therefore define the matrix

$$A = \begin{bmatrix} 0.6 & 0 & 0.2 \\ 0.4 & 0.6 & 0.2 \\ 0 & 0.4 & 0.6 \end{bmatrix}$$

and note that $\mathbf{x}_{k+1} = A\mathbf{x}_k$.

If we consider the first column of A, we see that the entries represent the percentages of party P's voters in the last election who vote for each of the three parties in the next election. Since everyone who voted for party P previously votes for one of the three parties in the next election, the sum of these percentages must be 1. This is true for each of the columns of A, which explains why A is a stochastic matrix.

We begin with the vector $\mathbf{x}_0 = \begin{pmatrix} 0.4 \\ 0.3 \\ 0.3 \end{pmatrix}$, the entries of which represent the percentage of voters voting for each of the three parties. Since every voter votes for one of the three parties, the sum of these entries must be 1, which means that $\mathbf{x}_0$ is a probability vector. We then find that

$$\mathbf{x}_1 = \begin{pmatrix} 0.300 \\ 0.400 \\ 0.300 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 0.240 \\ 0.420 \\ 0.340 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 0.212 \\ 0.416 \\ 0.372 \end{pmatrix}, \quad \ldots,$$

$$\mathbf{x}_5 = \begin{pmatrix} 0.199 \\ 0.404 \\ 0.397 \end{pmatrix}, \qquad \ldots, \qquad \mathbf{x}_{10} = \begin{pmatrix} 0.200 \\ 0.400 \\ 0.400 \end{pmatrix}, \quad \ldots$$

.

Notice that the vectors $\mathbf{x}_k$ are also probability vectors and that the sequence $\mathbf{x}_k$ seems to be converging to $\begin{pmatrix} 0.2 \\ 0.4 \\ 0.4 \end{pmatrix}$. It is this behavior that we would like to understand more fully by investigating the eigenvalues and eigenvectors of A.

We find that the eigenvalues of A are

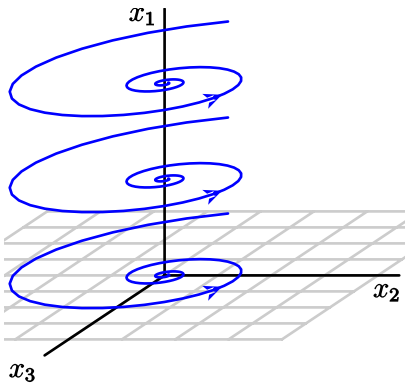
$$\lambda_1 = 1, \qquad \lambda_2 = 0.4 + 0.2i, \qquad \lambda_3 = 0.4 - 0.2i.$$

Notice that if $\mathbf{v}$ is an eigenvector of A with associated eigenvalue $\lambda_1 = 1$, then $A\mathbf{v} = 1\mathbf{v} = \mathbf{v}$. That is, $\mathbf{v}$ is unchanged when we multiply it by A.

Otherwise, we have $A = PEP^{-1}$ where

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.4 & -0.2 \\ 0 & 0.2 & 0.4 \end{bmatrix}$$

Notice that $|\lambda_2| = |\lambda_3| < 1$ so the trajectories $\mathbf{x}_k$ spiral into the eigenspace E_1 as indicated in the figure.



This tells us that the sequence $\mathbf{x}_k$ converges to a vector in E_1. It is straightforward to see that $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$ is a basis vector for E_1 because $A\mathbf{v} = \mathbf{v}$ so we expect that $\mathbf{x}_k$ will converge to a scalar multiple of $\mathbf{v}$. Indeed, since the vectors $\mathbf{x}_k$ are probability vectors, we expect them to converge to a probability vector in E_1.

We can find the probability vector in E_1 by finding the appropriate scalar multiple of $\mathbf{v}$. Notice that $c\mathbf{v} = \begin{pmatrix} c \\ 2c \\ 2c \end{pmatrix}$ is a probability vector when $c + 2c + 2c = 5c = 1$, which implies that $c = 1/5$. Therefore, $\mathbf{q} = \begin{pmatrix} 0.2 \\ 0.4 \\ 0.4 \end{pmatrix}$ is the unique probability vector in E_1. Since the sequence $\mathbf{x}_k$ converges to a probability vector in E_1, we see that $\mathbf{x}_k$ converges to $\mathbf{q}$, which agrees with the computations we showed above.

The role of the eigenvalues is important in this example. Since $\lambda_1 = 1$, we can find a probability vector $\mathbf{q}$ that is unchanged by multiplication by A. Also, the other eigenvalues satisfy $|\lambda_j| < 1$, which means that all the trajectories get pulled in to the eigenspace E_1. Since $\mathbf{x}_k$ is a sequence of probability vectors, these vectors converge to the probability vector $\mathbf{q}$ as they are pulled into E_1.

Markov chains

If we have a stochastic matrix A and a probability vector $\mathbf{x}_0$, we can form the sequence $\mathbf{x}_k$ where $\mathbf{x}_{k+1} = A\mathbf{x}_k$. We call this sequence of vectors a *Markov chain*. Exercise 4.5.5.6 explains why we can guarantee that the vectors $\mathbf{x}_k$ are probability vectors.

In the example that studied voting patterns, we constructed a Markov chain that described how the percentages of voters choosing different parties changed from one election to the next. We saw that the Markov chain converges to $\mathbf{q} = \begin{pmatrix} 0.2 \\ 0.4 \\ 0.4 \end{pmatrix}$, a probability vector in the eigenspace E_1. In other words, $\mathbf{q}$ is a probability vector that is unchanged under multiplication by A; that is, $A\mathbf{q} = \mathbf{q}$. This implies that, after a long time, 20% of voters choose party P, 40% choose Q, and 40% choose R.

Definition 4.5.2

If A is a stochastic matrix, we say that a probability vector $\mathbf{q}$ is a *steady-state* or *stationary* vector if $A\mathbf{q} = \mathbf{q}$.

An important question that arises from our previous example is

Question 4.5.3.

If A is a stochastic matrix and $\mathbf{x}_k$ a Markov chain, does $\mathbf{x}_k$ converge to a steady-state vector?

Activity 4.5.3.

Consider the matrices

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0.4 & 0.3 \\ 0.6 & 0.7 \end{bmatrix}.$$

1. Verify that both A and B are stochastic matrices.
2. Find the eigenvalues of A and then find a steady-state vector for A.
3. We will form the Markov chain beginning with the vector $\mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and defining $\mathbf{x}_{k+1} = A\mathbf{x}_k$. The Sage cell below constructs the first N terms of the Markov chain with the command `markov_chain(A, x0, N)`. Define the matrix A and vector $x0$ and evaluate the cell to find the first 10 terms of the Markov chain.
   What do you notice about the Markov chain? Does it converge to the steady-state vector for A?
4. Now find the eigenvalues of B along with a steady-state vector for B.
5. As before, find the first 10 terms in the Markov chain beginning with $\mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{x}_{k+1} = B\mathbf{x}_k$. What do you notice about the Markov chain? Does it converge to the steady-state vector for B?
6. What condition on the eigenvalues of a stochastic matrix will guarantee that a Markov chain will converge to a steady-state vector?

As this activity implies, the eigenvalues of a stochastic matrix tell us whether a Markov chain will converge to a steady-state vector. Here are a few important facts about the eigenvalues of a stochastic matrix.

- As is demonstrated in Exercise 4.5.5.8, $\lambda = 1$ is an eigenvalue of any stochastic matrix. We usually order the eigenvalues so it is the first eigenvalue meaning that $\lambda_1 = 1$.
- All other eigenvalues satisfy the property that $|\lambda_j| \leq 1$.
- Any stochastic matrix has at least one steady-state vector $\mathbf{q}$.

As illustrated in the activity, a Markov chain could fail to converge to a steady-state vector if $|\lambda_2| = 1$. This happens for the matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, whose eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = -1$.

However, if all but the first eigenvalue satisfy $|\lambda_j| < 1$, then there is a unique steady-state vector $\mathbf{q}$ and any Markov chain will converge to $\mathbf{q}$. This was the case for the matrix $B = \begin{bmatrix} 0.4 & 0.3 \\ 0.6 & 0.7 \end{bmatrix}$, whose eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 0.1$. In this case, any Markov chain will converge to the unique steady-state vector $\mathbf{q} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}$.

In this way, we see that the eigenvalues of a stochastic matrix tell us whether a Markov chain will converge to a steady-state vector. However, it is somewhat inconvenient to compute the eigenvalues to answer this question. Is there some way to conclude that every Markov chain will converge to a steady-state vector without actually computing the eigenvalues? It turns out that there is a simple condition on the matrix A that guarantees this.

Definition 4.5.4

We say that a matrix A is *positive* if either A or some power A^k has all positive entries.

Example 4.5.5

The matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is not positive. We can see this because some of the entries of A are zero and therefore not positive. In addition, we see that $A^2 = I$, $A^3 = A$ and so forth. Therefore, every power of A also has some zero entries, which means that A is not positive.

The matrix $B = \begin{bmatrix} 0.4 & 0.3 \\ 0.6 & 0.7 \end{bmatrix}$ is positive because every entry of B is positive.

Also, the matrix $C = \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix}$ clearly has a zero entry. However, $C^2 = \begin{bmatrix} 0.5 & 0.25 \\ 0.5 & 0.75 \end{bmatrix}$, which has all positive entries. Therefore, we see that C is a positive matrix.

Positive matrices are important because of the following theorem.

Theorem 4.5.6. Perron-Frobenius.

If A is a positive stochastic matrix, then the eigenvalues satisfy $\lambda_1 = 1$ and $|\lambda_j| < 1$ for $j > 1$. This means that A has a unique positive, steady-state vector $\mathbf{q}$ and that every Markov chain defined by A will converge to $\mathbf{q}$.

Activity 4.5.4.

We will explore the meaning of the Perron-Frobenius theorem in this activity.

1. Consider the matrix $C = \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix}$. This is a positive matrix, as we saw in the previous example. Find the eigenvectors of C and verify there is a unique steady-state vector.
2. Using the Sage cell below, construct the Markov chain with initial vector $\mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and describe what happens to $\mathbf{x}_k$ as k becomes large.
3. Construct another Markov chain with initial vector $\mathbf{x}_0 = \begin{pmatrix} 0.2 \\ 0.8 \end{pmatrix}$ and describe what happens to $\mathbf{x}_k$ as k becomes large.
4. Consider the matrix $D = \begin{bmatrix} 0 & 0.5 & 0 \\ 1 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and compute several powers of D below.
   Determine whether D is a positive matrix.
5. Find the eigenvalues of D and then find the steady-state vectors. Is there a unique steady-state vector?
6. What happens to the Markov chain defined by D with initial vector $\mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$? What happens to the Markov chain with initial vector $\mathbf{x}_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.
7. Explain how the matrices C and D, which we have considered in this activity, relate to the Perron-Frobenius theorem.

Google's PageRank algorithm

Markov chains and the Perron-Frobenius theorem are the central ingredients in Google's PageRank algorithm, developed by Google to assess the quality of web pages.

Suppose we enter “linear algebra” into Google's search engine. Google responds by telling us there are 24.9 million web pages containing those terms. On the first page, however, there are links to ten web pages that Google judges to have the highest quality and therefore the ones we are most likely to be interested in. How does Google assess the quality of web pages?

As the time this is being written, Google is tracking 30 trillion web pages. Clearly, this is too many for humans to evaluate. Plus, human evaluators may inject their own biases into their evaluations, perhaps even unintentionally. Google's idea is to use the structure of the Internet to assess the quality of web pages without any human intervention. For instance, if a web page has quality content, other web pages will link to it. This means that the number of links to a page reflect the quality of that page. In addition, we would expect a page to have even higher quality content if those links are coming from pages that are themselves assessed to have high quality. Simply said, if many quality pages link to a page, that page must itself be of high quality. This is

the essence of the PageRank algorithm, which we introduce in the next activity.

Activity 4.5.5.

We will consider a simple model of the Internet that has three pages and links between them as shown here. For instance, page 1 links to both pages 2 and 3, but page 2 only links to page 1.

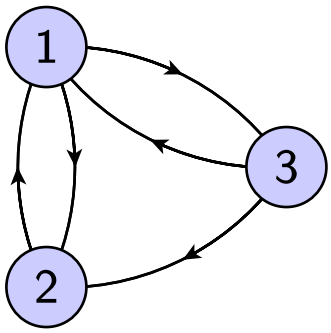


Figure 6.5.7: Our first Internet.

We will measure the quality of the jth page with a number x_j, which is called the PageRank of page j. The PageRank is determined by the following rule: each page divides its PageRank into equal pieces, one for each outgoing link, and gives one piece to each of the pages it links to. A page's PageRank is the sum of all the PageRank it receives from pages linking to it.

For instance, page 3 has two outgoing links. It therefore divides its PageRank x_3 in half and gives half to page 1. Page 2 has only one outgoing link so it gives all of its PageRank x_2 to page 1. We therefore have

$$x_1 = x_2 + \frac{1}{2}x_3.$$

1. Find similar expressions for x_2 and x_3.
2. We now form the PageRank vector $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$. Find a matrix G such that the expressions for x_1, x_2, and x_3 can be written in the form $G\mathbf{x} = \mathbf{x}$. The matrix G is called the "Google matrix".
3. Explain why G is a stochastic matrix.
4. Since $\mathbf{x}$ is defined by the equation $G\mathbf{x} = \mathbf{x}$, any vector in the eigenspace E_1 satisfies this equation. So that we might work with a specific vector, we will define the PageRank vector to be the steady-state vector of the stochastic matrix G. Find this steady state vector.
5. The PageRank vector $\mathbf{x}$ is composed of the PageRanks for each of the three pages. Which page of the three is assessed to have the highest quality? By referring to the structure of this small model of the Internet, explain why this is a good choice.
6. If we begin with the initial vector $\mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and form the Markov chain $\mathbf{x}_{k+1} = G\mathbf{x}_k$, what does the Perron-Frobenius theorem tell us about the long-term behavior of the Markov chain?
7. Verify that this Markov chain converges to the steady-state PageRank vector.

This activity shows us two ways to find the PageRank vector. In the first, we determine a steady-state vector directly by finding a description of the eigenspace E_1 and then finding the appropriate scalar multiple of a basis vector that gives us the steady-state vector. To find a description of the eigenspace E_1, however, we need to find the null space $Nul(G - I)$. Remember that the real Internet has 30 trillion pages so finding $Nul(G - I)$ requires us to row reduce a matrix with 30 trillion rows and columns. As we saw in Subsection 1.3, that is not computationally feasible.

As suggested by the activity, the second way to find the PageRank vector is to use a Markov chain that converges to the PageRank vector. Since multiplying a vector by a matrix is significantly less work than row reducing the matrix, this approach is

computationally feasible, and it is, in fact, how Google computes the PageRank vector.

Activity 4.5.6.

Consider the Internet with eight web pages, shown in Figure 4.5.8.

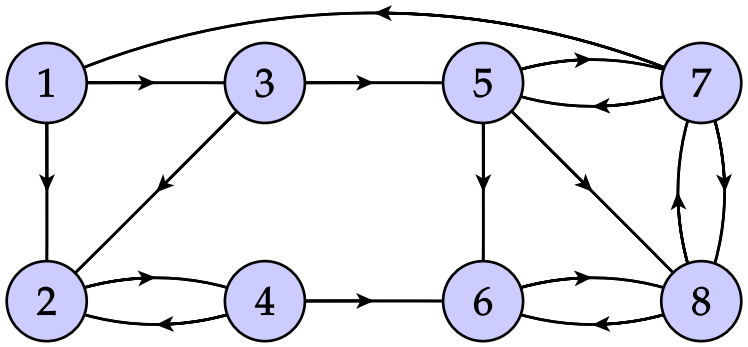


Figure 6.5.8: A simple model of the Internet with eight web pages.

1. Construct the Google matrix G for this Internet. Then use a Markov chain to find the steady-state PageRank vector $\mathbf{x}$.
2. What does this vector tell us about the relative quality of the pages in this Internet? Which page has the highest quality and which the lowest?
3. Now consider the Internet with five pages, shown in Figure 4.5.9.

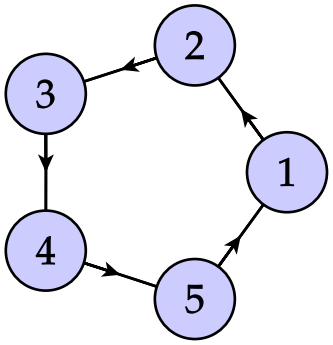


Figure 6.5.9: A model of the Internet with five web pages.

What happens when you begin the Markov chain with the vector $\mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$? Explain why this behavior is consistent with the Perron-Frobenius theorem.

4. What do you think the PageRank vector for this Internet should be? Is any one page of a higher quality than another?
5. Now consider the Internet with eight web pages, shown in Figure 4.5.10.

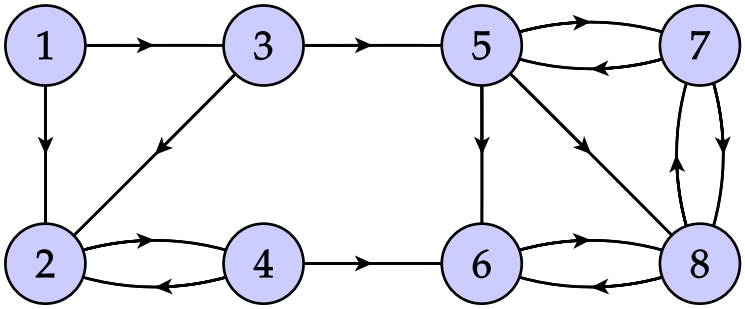


Figure 6.5.1: Another model of the Internet with eight web pages.

Notice that this version of the Internet is identical to the first one that we saw in this activity, except that a single link from page 7 to page 1 has been removed. We can therefore find its Google matrix G by slightly modifying the earlier matrix.

What is the long-term behavior of a Markov chain defined by G and why is this behavior not desirable? How is this behavior consistent with the Perron-Frobenius theorem?

The Perron-Frobenius theorem Theorem 4.5.6 tells us that a Markov chain $\mathbf{x}_{k+1} = G\mathbf{x}_k$ converges to a unique steady-state vector when the matrix G is positive. This means that G or some power of G should have only positive entries. Clearly, this is not the case for the matrix formed from the Internet in Figure 4.5.9.

We can understand the problem with the Internet shown in Figure 4.5.10 by adding a box around some of the pages as shown in Figure 4.5.11. Here we see that the pages outside of the box give up all of their PageRank to the pages inside the box. This is not desirable because the PageRanks of the pages outside of the box are found to be zero. Once again, the Google matrix G is not a positive matrix.

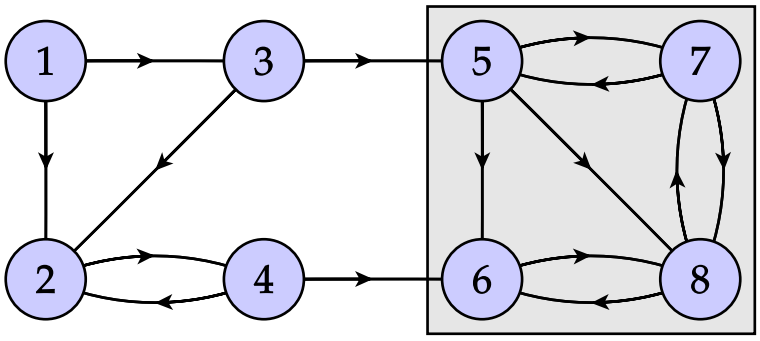


Figure 6.5.11: The pages outside the box give up all of their PageRank to the pages inside the box.

Google solves this problem by slightly modifying the Google matrix G to obtain a positive matrix G'. To understand this, think of the entries in the Google matrix as giving the probabilty that an Internet user follows a link from one page of another. To create a positive matrix, we will allow that user to randomly jump to any other page on the Internet with a small probability.

To make sense of this, suppose that there are N pages on our internet. The matrix

$$H_n = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix}$$

is a positive stochastic matrix describing a process where we can move from any page to another with equal probability. To

form the modified Google matrix G', we choose a parameter α that is used to mix G and H_n together; that is, G' is the positive stochastic matrix

$$G' = \alpha G + (1-\alpha)H_n.$$

In practice, it is thought that Google uses a value of $\alpha = 0.85$ (Google doesn't publish this number as it is a trade secret) so that we have

$$G' = 0.85G + 0.15H_n.$$

Intuitively, this means that an Internet user will randomly follow a link from one page to another 85% of the time and will randomly jump to any other page on the Internet 15% of the time. Since the matrix G' is positive, the Perron-Frobenius theorem tells us that any Markov chain will converge to a unique steady-state vector that we call the PageRank vector.

Activity 4.5.7.

The following Sage cell will generate the Markov chain for the modified Google matrix G if you simply enter the original Google matrix G in the appropriate line.

1. Consider the original Internet with three pages shown in Figure 4.5.7 and find the PageRank vector $\mathbf{x}$ using the modified Google matrix in the Sage cell above. How does this modified PageRank vector compare to the vector we found using the original Google matrix G?
2. Find the modified PageRank vector for the Internet shown in Figure 4.5.9. Explain why this vector seems to be the correct one.
3. Find the modified PageRank vector for the Internet shown in Figure 4.5.10. Explain why this modified PageRank vector fixes the problem that appeared with the original PageRank vector.

The ability to access almost anything we want to know through the Internet is something we take for granted in today's society. Without Google's PageRank algorithm, however, the Internet would be a chaotic place indeed; imagine trying to find a useful web page among the 30 trillion available pages without it. (There are, of course, other search algorithms, but Google's is the most widely used.) The fundamental role that Markov chains and the Perron-Frobenius theorem play in Google's algorithm demonstrates the vast power that mathematics has to shape our society.

Summary

This section explored stochastic matrices and Markov chains.

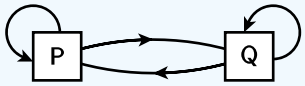
- A probability vector is one whose entries are nonnegative and whose columns add to 1. A stochastic matrix is a square matrix whose columns are probability vectors.
- A Markov chain is formed from a stochastic matrix A and an initial probability vector $\mathbf{x}_0$ using the rule $\mathbf{x}_{k+1} = A\mathbf{x}_k$. We may think of the sequence $\mathbf{x}_k$ as describing the evolution of some conserved quantity, such as the number of rental cars or voters, among a number of possible states over time.
- A steady-state vector $\mathbf{q}$ for a stochastic matrix A is a probability vector that satisfies $A\mathbf{q} = \mathbf{q}$.
- The Perron-Frobenius theorem tells us that, if A is a positive stochastic matrix, then every Markov chain defined by A converges to a unique, positive steady-state vector.
- Google's PageRank algorithm uses Markov chains and the Perron-Frobenius theorem to assess the relative quality of web pages on the Internet.

Exercises 4.5.5Exercises

1

Consider the following 2×2 stochastic matrices.

For each, make a copy of the diagram and label each edge to indicate the probability of that transition. Then find all the steady-state vectors and describe what happens to a Markov chain defined by that matrix.



1. $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$.
2. $\begin{bmatrix} 0.8 & 1 \\ 0.2 & 0 \end{bmatrix}$.
3. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
4. $\begin{bmatrix} 0.7 & 0.6 \\ 0.3 & 0.4 \end{bmatrix}$.

2

Every year, people move between urban (U), suburban (S), and rural (R) populations with the probabilities given in Figure 4.5.12.

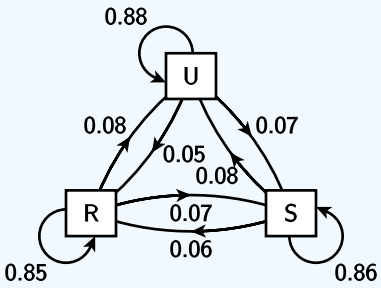


Figure 6.5.12: The flow between urban, suburban, and rural populations.

1. Construct the stochastic matrix A describing the movement of people.
2. Explain what the Perron-Frobenius theorem tells us about the existence of a steady-state vector $\mathbf{q}$ and the behavior of a Markov chain.
3. Use the Sage cell below to find the some terms of a Markov chain.
4. Describe the long-term distribution of people among urban, suburban, and rural populations.

3

Determine whether the following statements are true or false and provide a justification of your response.

1. Every stochastic matrix has a steady-state vector.
2. If A is a stochastic matrix, then any Markov chain defined by A converges to a steady-state vector.
3. If A is a stochastic matrix, then $\lambda = 1$ is an eigenvalue and all the other eigenvalues satisfy $|\lambda| < 1$.
4. A positive stochastic matrix has a unique steady-state vector.
5. If A is an invertible stochastic matrix, then so is A^{-1}.

4

Consider the stochastic matrix

$$A = \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0 & 0.6 & 0.2 \\ 0 & 0.2 & 0.6 \end{bmatrix}.$$

1. Find the eigenvalues of A.
2. Do the conditions of the Perron-Frobenius theorem apply to this matrix?
3. Find the steady-state vectors of A.
4. What can we guarantee about the long-term behavior of a Markov chain defined by the matrix A?

5

Explain your responses to the following.

1. Why does Google use a Markov chain to compute the PageRank vector?
2. Describe two problems that can happen when Google constructs a Markov chain using the Google matrix G.
3. Describe how these problems are consistent with the Perron-Frobenius theorem.
4. Describe why the Perron-Frobenius theorem suggests creating a Markov chain using the modified Google matrix $G' = \alpha G + (1-\alpha)H_n$.

In the next few exercises, we will consider the $1 \times n$ matrix $S = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}$.

6

Suppose that A is a stochastic matrix and that $\mathbf{x}$ is a probability vector. We would like to explain why the product $A\mathbf{x}$ is a probability vector.

1. Explain why $\mathbf{x} = \begin{pmatrix} 0.4 \\ 0.5 \\ 0.1 \end{pmatrix}$ is a probability vector and then find the product $S\mathbf{x}$.
2. More generally, if $\mathbf{x}$ is any probability vector, what is the product $S\mathbf{x}$?
3. If A is a stochastic matrix, explain why $SA = S$.
4. Explain why $A\mathbf{x}$ is a probability vector by considering the product $SA\mathbf{x}$.

7

Using the results of the previous exercise, we would like to explain why A^2 is a stochastic matrix if A is stochastic.

1. Suppose that A and B are stochastic matrices. Explain why the product AB is a stochastic matrix by considering the product SAB.
2. Explain why A^2 is a stochastic matrix.
3. How do the steady-state vectors of A^2 compare to the steady-state vectors of A?

8

This exercise explains why $\lambda = 1$ is an eigenvalue of a stochastic matrix A. To conclude that $\lambda = 1$ is an eigenvalue, we need to know that $A - I$ is not invertible.

1. What is the product $S(A - I)$?

2. Explain why $Col(A-I)$ is contained in $Nul(S)$.
3. What is the product $S\mathbf{e}_1$?
4. Explain why $\mathbf{e}_1$ is not contained in the column space $Col(A-I)$.
5. Explain why we can conclude that $A-I$ is not invertible and that $\lambda=1$ is an eigenvalue of A.

9

We saw a couple of model Internets in which a Markov chain defined by the Google matrix G did not converge to an appropriate PageRank vector. For this reason, Google defines the matrix

$$H_n = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix},$$

where n is the number of web pages, and constructs a Markov chain from the modified Google matrix

$$G' = \alpha G + (1-\alpha)H_n.$$

Since G' is positive, the Markov chain is guaranteed to converge to a unique steady-state vector.

We said that Google chooses $\alpha = 0.85$ so we might wonder why this is a good choice. We will explore the role of α in this exercise. Let's consider the model Internet described in Figure 4.5.9 and construct the Google matrix G. In the Sage cell below, you can enter the matrix G and choose a value for α.

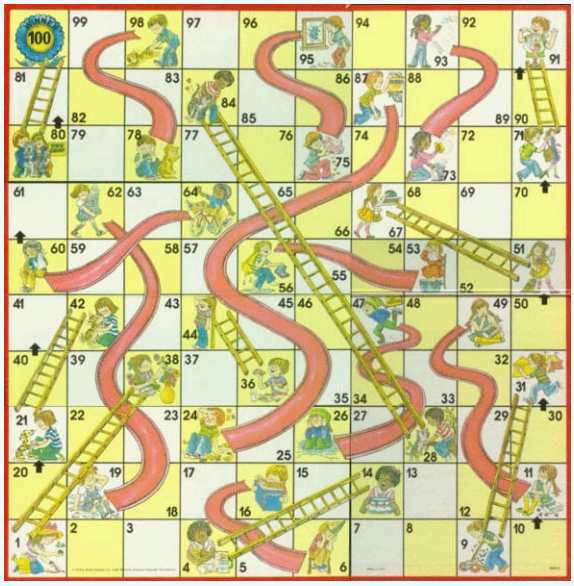
1. Let's begin with $\alpha = 0$. With this choice, what is the matrix $G' = \alpha G + (1-\alpha)H_n$? Construct a Markov chain using the Sage cell above. How many steps are required for the Markov chain to converge to the accuracy with which the vectors $\mathbf{x}_k$ are displayed?
2. Now choose $\alpha = 0.25$. How many steps are required for the Markov chain to converge to the accuracy at which the vectors $\mathbf{x}_k$ are displayed?
3. Repeat this experiment with $\alpha = 0.5$ and $\alpha = 0.75$.
4. What happens if $\alpha = 1$?

This experiment gives some insight into the choice of α. The smaller α is, the faster the Markov chain converges. This is important; since the matrix G' that Google works with is so large, we would like to minimize the number of terms in the Markov chain that we need to compute. On the other hand, as we lower α, the matrix $G' = \alpha G + (1-\alpha)H_n$ begins to resemble H_n more and G less. The value $\alpha = 0.85$ is chosen so that the matrix G' sufficiently resembles G while having the Markov chain converge in a reasonable amount of steps.

10

This exercise will analyze the board game *Chutes and Ladders*, or at least a simplified version of it.

The board for this game consists of 100 squares arranged in a 10×10 grid and numbered 1 to 100. There are pairs of squares joined by a ladder and pairs joined by a chute. All players begin in square 1 and take turns rolling a die. On their turn, a player will move ahead the number of squares indicated on the die. If they arrive at a square at the bottom of a ladder, they move to the square at the top of the ladder. If they arrive at a square at the top of a chute, the move down to the square at the bottom of the chute. The winner is the first player to reach square 100.



1. We begin by playing a simpler version of this game with only eight squares laid out in a row as shown in Figure 4.5.13 and containing neither chutes nor ladders. Rather than a six-sided die, we will flip a coin and move ahead one or two squares depending on the result whether we have heads or tails. If we are on square 7, we move ahead to square 8 regardless of the coin flip, and if we are on square 8, we will stay there forever.

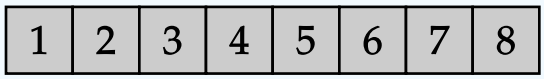


Figure 6.5.13: A simple version of Chutes and Ladders with neither chutes nor ladders

2. Construct the 8×8 matrix A that records the probability that a player moves from one square to another on one move. For instance, if a player is on square 2, there is a 50% chance they are on square 3 and a 50% chance they are on square 4 at the end of the next move.
3. Since we begin the game on square 1, the initial vector $\mathbf{x}_0 = \mathbf{e}_1$. Generate a few terms of the Markov chain $\mathbf{x}_{k+1} = A\mathbf{x}_k$.
4. What is the probability that we arrive at square 8 by the fourth move? By the sixth move? By the seventh move?
5. We will now modify the game by adding one chute and one ladder as shown in Figure 4.5.14.

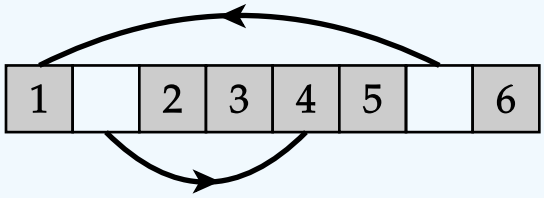


Figure 6.5.14: A version of Chutes and Ladders with one chute and one ladder.

Even though there are eight squares, we only need to consider six of them. For instance, if we arrive at the first white square, we move up to square 4. Similarly, if we arrive at the second white square, we move down to square 1.

Once again, construct the 6×6 stochastic matrix that records the probability that we move from one square to another on a given turn and generate some terms in the Markov chain that begins with $\mathbf{x}_0 = \mathbf{e}_1$.

1. What is the smallest number of moves we can make and arrive at square 6? What is the probability that we arrive at square 6 using this number of moves?
2. What is the probability that we arrive at square 6 after five moves?

3. What is the probability that we are still on square 1 after five moves? After seven moves? After nine moves?
4. After how many moves do we have a 90% chance of having arrived at square 6?
5. Find the steady-state vector and discuss what this vector implies about the game.

One can analyze the full version of Chutes and Ladders having 100 squares in the same way. Without any chutes or ladders, one finds that the average number of moves required to reach square 100 is 29.0. Once we add the chutes and ladders back in, the average number of moves required to reach square 100 is 27.1. This shows that the average number of moves does not change significantly when we add the chutes and ladders. There is, however, much more variation in the possibilities because it is possible to reach square 100 much more quickly and much more slowly.

This page titled 6.5: Markov chains and Google's PageRank algorithm is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by David Austin via source content that was edited to the style and standards of the LibreTexts platform.

- 4.5: Markov chains and Google's PageRank algorithm by David Austin is licensed CC BY 4.0. Original source: https://davidaustinm.github.io/ula/ula.html.

7: HONORS PROJECTS

7.1: Stochastic Matrices

7.2: An Application to Markov Chains

7.2.1: Absorbing Markov Chains

7.2.1.1: Absorbing Markov Chains (Exercises)

7.2.2: Introduction to Markov Chains

7.2.2.1: Introduction to Markov Chains (Exercises)

7.2.3: Regular Markov Chains

7.2.3.1: Regular Markov Chains (Exercises)

7.3: The Method of Least Squares

7.4: An Application to Linear Recurrences

7.4E: An Application to Linear Recurrences Exercises

7.5: An Application to Systems of Differential Equations

7.5E: An Application to Systems of Differential Equations Exercises

7.6: An Application to Computer Graphics

7.6E: An Application to Computer Graphics Exercises

7.7: An Application to Input-Output Economic Models

7.7E: An Application to Input-Output Economic Models Exercises

7.8: The Complex Number System

7.8.1: Solve Systems of Linear Equations with Complex Coefficients

This page titled 7: Honors Projects is shared under a license and was authored, remixed, and/or curated by LibreTexts.

7.1: 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 7.1.1: 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 7.1.1

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):
>
> $$\begin{gathered} 30\% \text{ of the movies from kiosk } 1 \\ 40\% \text{ of the movies from kiosk } 2 \\ 30\% \text{ of the movies from kiosk } 3 \end{gathered}$$

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 7.1.2: 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:

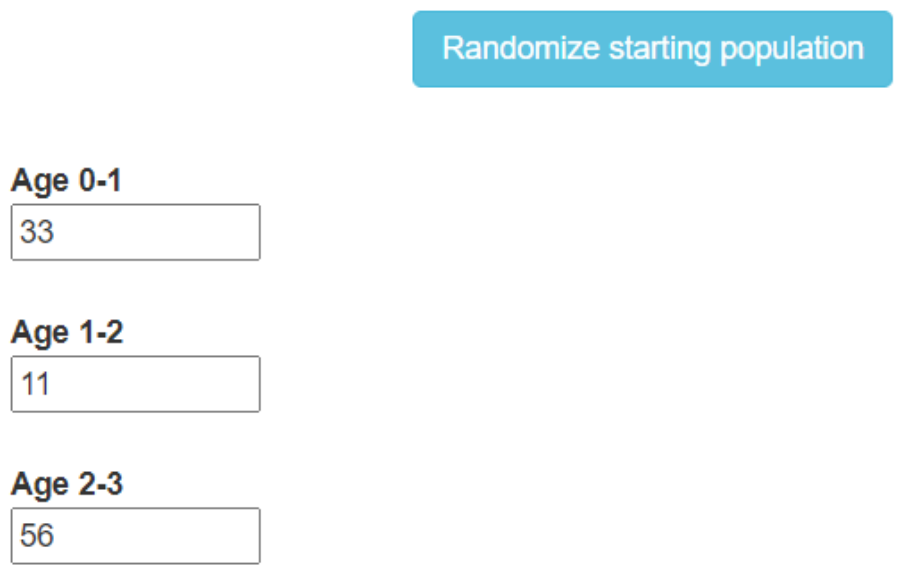


Figure 7.1.1: Left: the population of rabbits in a given year. Right: the proportions of rabbits in that year. Choose any values you like for the starting population, and click “Advance 1 year” several times. Notice that the ratio $x_t : y_t : z_t$ approaches $16 : 4 : 1$, and that all three quantities eventually double each year.

Stochastic Matrices and the Steady State

In this subsection, we discuss difference equations representing *probabilities*, like the Red Box example 7.1.1. 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 7.1.2: 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 7.1.1

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 7.1.3

Continuing with Red Box example 7.1.1, 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,
>
> $$\begin{aligned} &30\% \text{ will be returned to kiosk } 1 \\ &30\% \text{ will be returned to kiosk } 2 \\ &40\% \text{ will be returned to kiosk } 3. \end{aligned}$$
>
> 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 7.1.2

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 7.1.1

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 available at the end of this section.

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 7.1.3: 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 7.1.1: 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 7.1.4: 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

$$A^tx = a_1u_1 + \frac{a_2}{2^t}u_2 \ \to\ a_1u_1$$

as $t \to \infty$. This shows that $A^t x$ approaches

$$a_1 u_1 = \begin{pmatrix} a_1 \\ a_1 \end{pmatrix}.$$

Note that the sum of the entries of $a_1 u_1$ is equal to the sum of the entries of $a_1 u_1 + a_2 u_2$, 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.

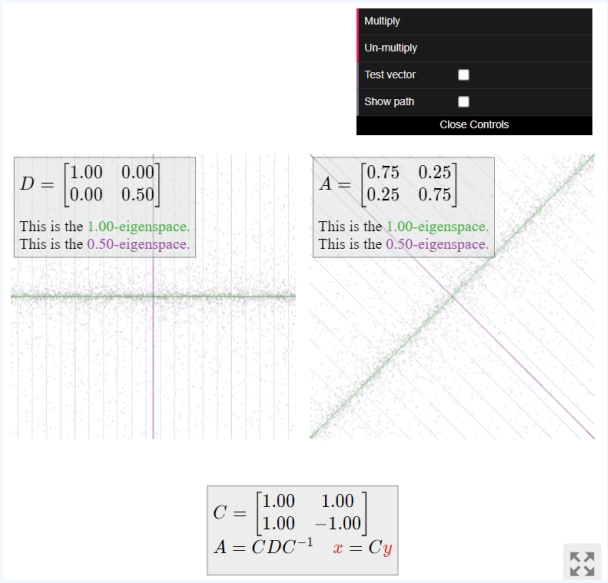


Figure 7.1.2: Dynamics of the stochastic matrix A. Click "multiply" to multiply the colored points by D on the left and A on the right. Note that on both sides, all vectors are "sucked into the 1-eigenspace" (the red line).

Example 7.1.5

Continuing with Red Box example 7.1.1, 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.

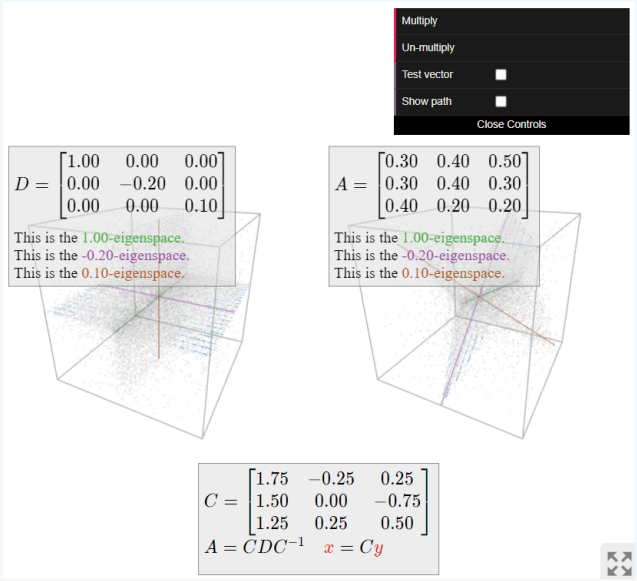


Figure 7.1.3: Dynamics of the stochastic matrix A. Click "multiply" to multiply the colored points by D on the left and A on the right. Note that on both sides, all vectors are "sucked into the 1-eigenspace" (the green line). (We have scaled C by $1/4$ so that vectors have roughly the same size on the 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 7.1.4: 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 7.1.5: 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 i th column contains the number $1/m$, m times, and the number zero in the other entries.

Example 7.1.6

Consider the following internet with only four pages. Links are indicated by arrows.

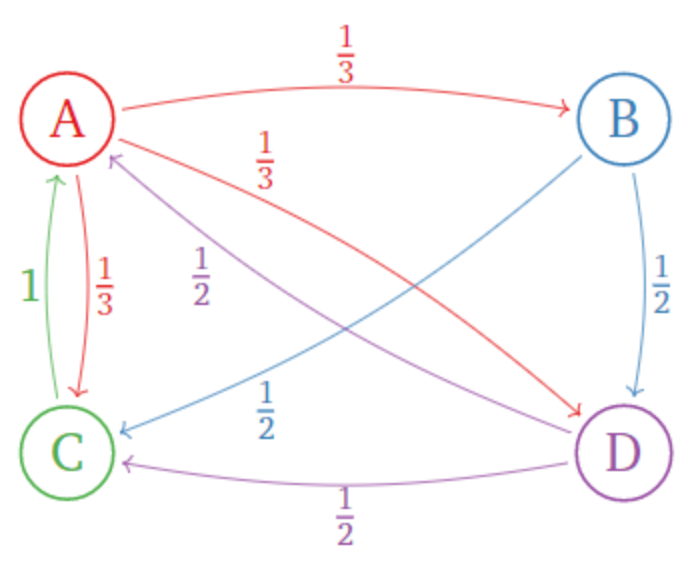


Figure 7.1.4

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 7.1.1

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 7.1.7: A page with no links

Consider the following internet with three pages:

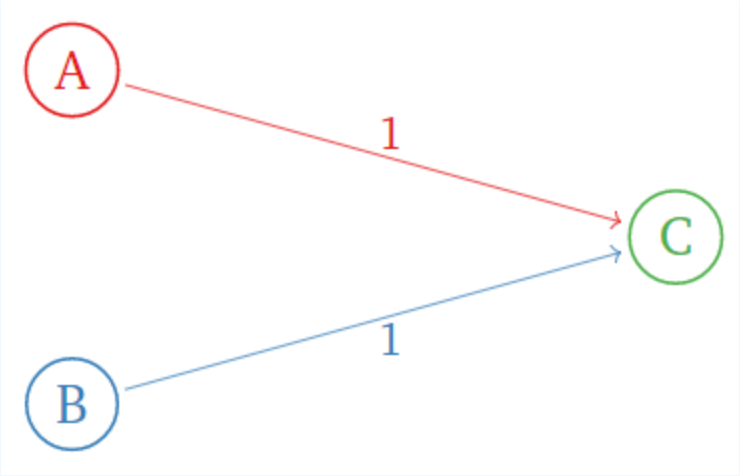


Figure 7.1.5

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 7.1.8: Disconnected Internet

Consider the following internet:



Figure 7.1.6

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 7.1.6: 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 7.1.2

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 7.1.3: 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 7.1.9: PageRank Vector

What is the PageRank vector for the following internet? (Use the damping factor $p = 0.15$.)

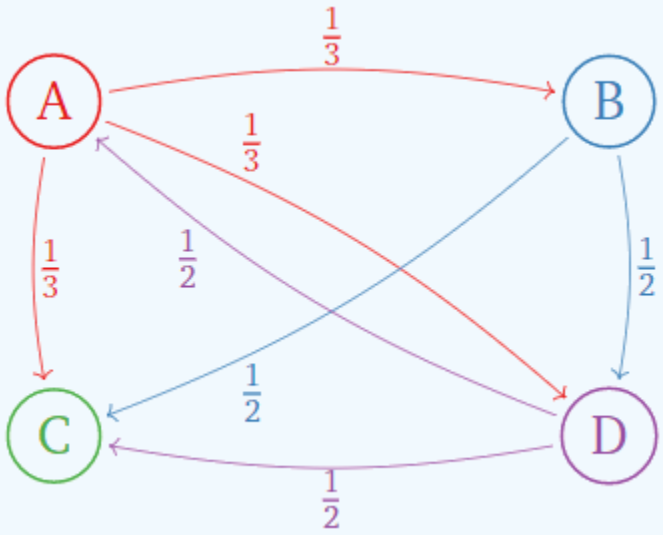


Figure 7.1.7

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:

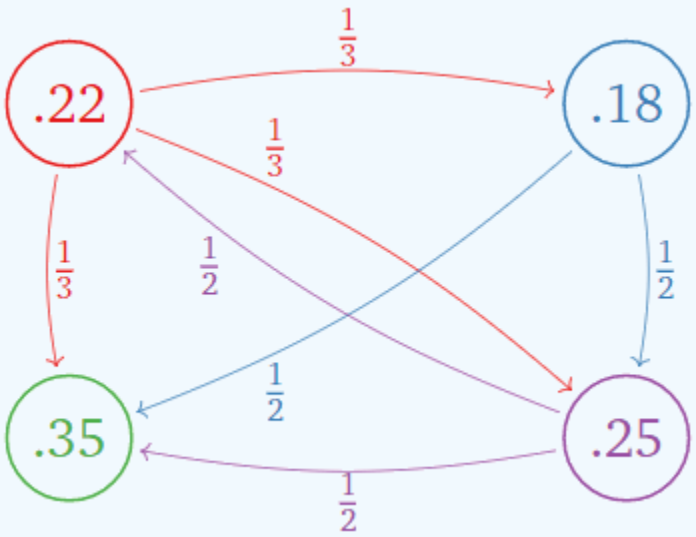


Figure 7.1.8

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.

Answers

Fact 7.1.1

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:

$$\det(A - \lambda I_n) = \det\left((A - \lambda I_n)^T\right) = \det(A^T - \lambda I_n).$$

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 $|\lambda| \leq 1$.

This page titled 7.1: Stochastic Matrices is shared under a GNU FDL license and was authored, remixed, and/or curated by Dan Margalit and Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- 5.6: Stochastic Matrices by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

7.2: An Application to Markov Chains

Many natural phenomena progress through various stages and can be in a variety of states at each stage. For example, the weather in a given city progresses day by day and, on any given day, may be sunny or rainy. Here the states are “sun” and “rain,” and the weather progresses from one state to another in daily stages. Another example might be a football team: The stages of its evolution are the games it plays, and the possible states are “win,” “draw,” and “loss.”

The general setup is as follows: A real conceptual “system” is run generating a sequence of outcomes. The system evolves through a series of “stages,” and at any stage it can be in any one of a finite number of “states.” At any given stage, the state to which it will go at the next stage depends on the past and present history of the system—that is, on the sequence of states it has occupied to date.

Definition: Markov Chain

A **Markov chain** is such an evolving system wherein the state to which it will go next depends only on its present state and does not depend on the earlier history of the system.

Even in the case of a Markov chain, the state the system will occupy at any stage is determined only in terms of probabilities. In other words, chance plays a role. For example, if a football team wins a particular game, we do not know whether it will win, draw, or lose the next game. On the other hand, we may know that the team tends to persist in winning streaks; for example, if it wins one game it may win the next game $\frac{1}{2}$ of the time, lose $\frac{4}{10}$ of the time, and draw $\frac{1}{10}$ of the time. These fractions are called the **probabilities** of these various possibilities. Similarly, if the team loses, it may lose the next game with probability $\frac{1}{2}$ (that is, half the time), win with probability $\frac{1}{4}$, and draw with probability $\frac{1}{4}$. The probabilities of the various outcomes after a drawn game will also be known.

We shall treat probabilities informally here: *The probability that a given event will occur is the long-run proportion of the time that the event does indeed occur*. Hence, all probabilities are numbers between 0 and 1. A probability of 0 means the event is impossible and never occurs; events with probability 1 are certain to occur.

If a Markov chain is in a particular state, the probabilities that it goes to the various states at the next stage of its evolution are called the **transition probabilities** for the chain, and they are assumed to be known quantities. To motivate the general conditions that follow, consider the following simple example. Here the system is a man, the stages are his successive lunches, and the states are the two restaurants he chooses.

Example 7.2.1

A man always eats lunch at one of two restaurants, A and B. He never eats at A twice in a row. However, if he eats at B, he is three times as likely to eat at B next time as at A. Initially, he is equally likely to eat at either restaurant.

1. What is the probability that he eats at A on the third day after the initial one?
2. What proportion of his lunches does he eat at A?

Solution

The table of transition probabilities follows. The A column indicates that if he eats at A on one day, he never eats there again on the next day and so is certain to go to B.

	Present Lunch A	Present Lunch B
Next Lunch A	0	0.25
Next Lunch B	1	0.75

The B column shows that, if he eats at B on one day, he will eat there on the next day $\frac{3}{4}$ of the time and switches to A only $\frac{1}{4}$ of the time.

The restaurant he visits on a given day is not determined. The most that we can expect is to know the probability that he will visit A or B on that day.

Let $\mathbf{s}_m = \begin{bmatrix} s_1^{(m)} \\ s_2^{(m)} \end{bmatrix}$ denote the *state vector* for day m. Here $s_1^{(m)}$ denotes the probability that he eats at A on day m, and $s_2^{(m)}$ is the probability that he eats at B on day m. It is convenient to let $\mathbf{s}_0$ correspond to the initial day. Because he is equally likely to eat at A or B on that initial day, $s_1^{(0)} = 0.5$ and $s_2^{(0)} = 0.5$, so $\mathbf{s}_0 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$. Now let

$$P = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix}$$

denote the *transition matrix*. We claim that the relationship

$$\mathbf{s}_{m+1} = P\mathbf{s}_m$$

holds for all integers $m \geq 0$. This will be derived later; for now, we use it as follows to successively compute $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \ldots$.

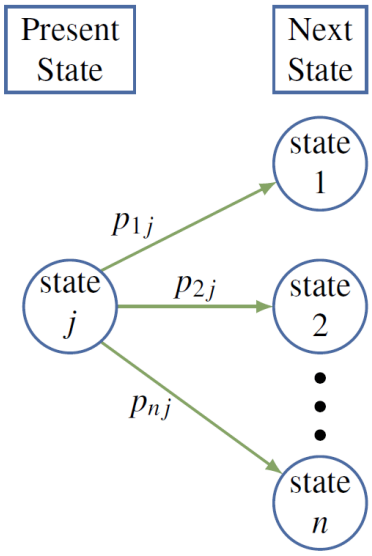
$$\mathbf{s}_1 = P\mathbf{s}_0 = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.125 \\ 0.875 \end{bmatrix}$$
$$\mathbf{s}_2 = P\mathbf{s}_1 = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix} \begin{bmatrix} 0.125 \\ 0.875 \end{bmatrix} = \begin{bmatrix} 0.21875 \\ 0.78125 \end{bmatrix}$$
$$\mathbf{s}_3 = P\mathbf{s}_2 = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix} \begin{bmatrix} 0.21875 \\ 0.78125 \end{bmatrix} = \begin{bmatrix} 0.1953125 \\ 0.8046875 \end{bmatrix}$$

Hence, the probability that his third lunch (after the initial one) is at A is approximately 0.195, whereas the probability that it is at B is 0.805. If we carry these calculations on, the next state vectors are (to five figures):

$$\mathbf{s}_4 = \begin{bmatrix} 0.20117 \\ 0.79883 \end{bmatrix} \quad \mathbf{s}_5 = \begin{bmatrix} 0.19971 \\ 0.80029 \end{bmatrix}$$
$$\mathbf{s}_6 = \begin{bmatrix} 0.20007 \\ 0.79993 \end{bmatrix} \quad \mathbf{s}_7 = \begin{bmatrix} 0.19998 \\ 0.80002 \end{bmatrix}$$

Moreover, as m increases the entries of $\mathbf{s}_m$ get closer and closer to the corresponding entries of $\begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$. Hence, in the long run, he eats 20% of his lunches at A and 80% at B.

Example 2.9.1 incorporates most of the essential features of all Markov chains. The general model is as follows: The system evolves through various stages and at each stage can be in exactly one of n distinct states. It progresses through a sequence of states as time goes on. If a Markov chain is in state j at a particular stage of its development, the probability p_{ij} that it goes to state i at the next stage is called the **transition probability**. The $n \times n$ matrix $P = [p_{ij}]$ is called the **transition matrix** for the Markov chain. The situation is depicted graphically in the diagram.



We make one important assumption about the transition matrix $P = [p_{ij}]$: It does *not* depend on which stage the process is in. This assumption means that the transition probabilities are *independent of time*—that is, they do not change as time goes on. It is this assumption that distinguishes Markov chains in the literature of this subject.

Example 7.2.2

Suppose the transition matrix of a three-state Markov chain is

	R_1	R_2	R_3
R_1	0.5	0.25	0.25
R_2	1	0.5	0.25
R_3	0.5	0.25	0.5

If, for example, the system is in state 2, then column 2 lists the probabilities of where it goes next. Thus, the probability is $p_{12} = 0.1$ that it goes from state 2 to state 1, and the probability is $p_{22} = 0.9$ that it goes from state 2 to state 2. The fact that $p_{32} = 0$ means that it is impossible for it to go from state 2 to state 3 at the next stage.

Consider the jth column of the transition matrix P.

$$\begin{bmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{nj} \end{bmatrix}$$

If the system is in state j at some stage of its evolution, the transition probabilities $p_{1j}, p_{2j}, \ldots, p_{nj}$ represent the fraction of the time that the system will move to state 1, state 2, ..., state n, respectively, at the next stage. We assume that it has to go to *some* state at each transition, so the sum of these probabilities is 1:

$$p_{1j} + p_{2j} + \cdots + p_{nj} = 1 \quad \text{for each } j$$

Thus, the columns of P all sum to 1 and the entries of P lie between 0 and 1. Hence P is called a **stochastic matrix**.

As in Example 2.9.1, we introduce the following notation: Let $s_i^{(m)}$ denote the probability that the system is in state i after m transitions. The $n \times 1$ matrices

$$\mathbf{s}_m = \begin{bmatrix} s_1^{(m)} \\ s_2^{(m)} \\ \vdots \\ s_n^{(m)} \end{bmatrix} \quad m = 0, 1, 2, \ldots$$

are called the **state vectors** for the Markov chain. Note that the sum of the entries of $\mathbf{s}_m$ must equal 1 because the system must be in *some* state after m transitions. The matrix $\mathbf{s}_0$ is called the **initial state vector** for the Markov chain and is given as part of the data of the particular chain. For example, if the chain has only two states, then an initial vector $\mathbf{s}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ means that it started in state 1. If it started in state 2, the initial vector would be $\mathbf{s}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. If $\mathbf{s}_0 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$, it is equally likely that the system started in state 1 or in state 2.

Theorem 7.2.1

Let P be the transition matrix for an n-state Markov chain. If $\mathbf{s}_m$ is the state vector at stage m, then

$$\mathbf{s}_{m+1} = P\mathbf{s}_m$$

for each $m = 0, 1, 2, \ldots$.

Heuristic Proof

Suppose that the Markov chain has been run N times, each time starting with the same initial state vector. Recall that p_{ij} is the proportion of the time the system goes from state j at some stage to state i at the next stage, whereas $s_i^{(m)}$ is the proportion of the time it is in state i at stage m. Hence

$$s_i^{m+1}N$$

is (approximately) the number of times the system is in state i at stage $m + 1$. We are going to calculate this number another way. The system got to state i at stage $m + 1$ through *some* other state (say state j) at stage m. The number of times it was *in* state j at that stage is (approximately) $s_j^{(m)}N$, so the number of times it got to state i via state j is $p_{ij}(s_j^{(m)}N)$. Summing over j gives the number of times the system is in state i (at stage $m + 1$). This is the number we calculated before, so

$$s_i^{(m+1)}N = p_{i1}s_1^{(m)}N + p_{i2}s_2^{(m)}N + \cdots + p_{in}s_n^{(m)}N$$

Dividing by N gives $s_i^{(m+1)} = p_{i1}s_1^{(m)} + p_{i2}s_2^{(m)} + \cdots + p_{in}s_n^{(m)}$ for each i, and this can be expressed as the matrix equation $\mathbf{s}_{m+1} = P\mathbf{s}_m$.

□

If the initial probability vector $\mathbf{s}_0$ and the transition matrix P are given, Theorem 2.9.1 gives $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \ldots$, one after the other, as follows:

$$\begin{aligned} \mathbf{s}_1 &= P\mathbf{s}_0 \\ \mathbf{s}_2 &= P\mathbf{s}_1 \\ \mathbf{s}_3 &= P\mathbf{s}_2 \\ &\vdots \end{aligned}$$

Hence, the state vector $\mathbf{s}_m$ is completely determined for each $m = 0, 1, 2, \ldots$ by P and $\mathbf{s}_0$.

Example 7.2.3

A wolf pack always hunts in one of three regions R_1, R_2, and R_3. Its hunting habits are as follows:

1. If it hunts in some region one day, it is as likely as not to hunt there again the next day.
2. If it hunts in R_1, it never hunts in R_2 the next day.
3. If it hunts in R_2 or R_3, it is equally likely to hunt in each of the other regions the next day.

If the pack hunts in R_1 on Monday, find the probability that it hunts there on Thursday.

Solution

The stages of this process are the successive days; the states are the three regions. The transition matrix P is determined as follows (see the table): The first habit asserts that $p_{11} = p_{22} = p_{33} = \frac{1}{2}$. Now column 1 displays what happens when the pack starts in R_1: It never goes to state 2, so $p_{21} = 0$ and, because the column must sum to 1, $p_{31} = \frac{1}{2}$. Column 2 describes what happens if it starts in R_2: $p_{22} = \frac{1}{2}$ and p_{12} and p_{32} are equal (by habit 3), so $p_{12} = p_{32} = \frac{1}{2}$ because the column sum must equal 1. Column 3 is filled in a similar way.

	R_1	R_2	R_3
R_1	1/2	1/4	1/4
R_2	0	1/2	1/4
R_3	1/2	1/4	1/2

Now let Monday be the initial stage. Then $\mathrm{s}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ because the pack hunts in R_1 on that day. Then s_1, s_2, and s_3 describe Tuesday, Wednesday, and Thursday, respectively, and we compute them using Theorem 2.9.1.

$$\mathrm{s}_1 = P\mathrm{s}_0 = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix} \quad \mathrm{s}_2 = P\mathrm{s}_1 = \begin{bmatrix} \frac{3}{8} \\ \frac{1}{8} \\ \frac{4}{8} \end{bmatrix} \quad \mathrm{s}_3 = P\mathrm{s}_2 = \begin{bmatrix} \frac{11}{32} \\ \frac{6}{32} \\ \frac{15}{32} \end{bmatrix}$$

Hence, the probability that the pack hunts in Region R_1 on Thursday is $\frac{11}{32}$.

Steady State Vector

Another phenomenon that was observed in Example 2.9.1 can be expressed in general terms. The state vectors $\mathrm{s}_0, \mathrm{s}_1, \mathrm{s}_2, \ldots$ were calculated in that example and were found to "approach" $\mathrm{s} = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$. This means that the first component of s_m becomes and remains very close to 0.2 as m becomes large, whereas the second component gets close to 0.8 as m increases. When this is the case, we say that s_m **converges** to s. For large m, then, there is very little error in taking $\mathrm{s}_m = \mathrm{s}$, so the long-term probability that the system is in state 1 is 0.2, whereas the probability that it is in state 2 is 0.8. In Example 2.9.1, enough state vectors were computed for the limiting vector s to be apparent. However, there is a better way to do this that works in most cases.

Suppose P is the transition matrix of a Markov chain, and assume that the state vectors s_m converge to a limiting vector s. Then s_m is very close to s for sufficiently large m, so s_{m+1} is also very close to s. Thus, the equation $\mathrm{s}_{m+1} = P\mathrm{s}_m$ from Theorem 2.9.1 is closely approximated by

$$\mathrm{s} = P\mathrm{s}$$

so it is not surprising that s should be a solution to this matrix equation. Moreover, it is easily solved because it can be written as a system of homogeneous linear equations

$$(I - P)\mathrm{s} = \mathbf{0}$$

with the entries of s as variables.

In Example 2.9.1, where $P = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix}$, the general solution to $(I - P)\mathrm{s} = \mathbf{0}$ is $\mathrm{s} = \begin{bmatrix} t \\ 4t \end{bmatrix}$, where t is a parameter. But if we insist that the entries of S sum to 1 (as must be true of all state vectors), we find $t = 0.2$ and so $\mathrm{s} = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$ as before.

All this is predicated on the existence of a limiting vector for the sequence of state vectors of the Markov chain, and such a vector may not always exist. However, it does exist in one commonly occurring situation. A stochastic matrix P is called **regular** if some power P^m of P has every entry greater than zero. The matrix $P = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix}$ of Example 2.9.1 is regular (in this case, each entry of P^2 is positive), and the general theorem is as follows:

Theorem 7.2.2

Let P be the transition matrix of a Markov chain and assume that P is regular. Then there is a unique column matrix s satisfying the following conditions:

1. $P\mathrm{s} = \mathrm{s}$.
2. The entries of s are positive and sum to 1.

Moreover, condition 1 can be written as

$$(I - P)\mathrm{s} = \mathbf{0}$$

and so gives a homogeneous system of linear equations for s. Finally, the sequence of state vectors $\mathrm{s}_0, \mathrm{s}_1, \mathrm{s}_2, \ldots$ converges to s in the sense that if m is large enough, each entry of s_m is closely approximated by the corresponding entry of s.

This theorem will not be proved here.[1]

If P is the regular transition matrix of a Markov chain, the column s satisfying conditions 1 and 2 of Theorem 2.9.2 is called the **steady-state vector** for the Markov chain. The entries of s are the long-term probabilities that the chain will be in each of the various states.

Example 7.2.4

A man eats one of three soups—beef, chicken, and vegetable—each day. He never eats the same soup two days in a row. If he eats beef soup on a certain day, he is equally likely to eat each of the others the next day; if he does not eat beef soup, he is twice as likely to eat it the next day as the alternative.

1. If he has beef soup one day, what is the probability that he has it again two days later?
2. What are the long-run probabilities that he eats each of the three soups?

Solution

The states here are B, C, and V, the three soups. The transition matrix P is given in the table. (Recall that, for each state, the corresponding column lists the probabilities for the next state.)

	B	C	V
B	0	2/3	2/3
C	1/2	0	0.25
V	1/2	1/3	0

If he has beef soup initially, then the initial state vector is

$$\mathbf{s}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Then two days later the state vector is $\mathbf{s}_2$. If P is the transition matrix, then

$$\mathbf{s}_1 = P\mathbf{s}_0 = \frac{1}{2}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{s}_2 = P\mathbf{s}_1 = \frac{1}{6}\begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix}$$

so he eats beef soup two days later with probability $\frac{2}{3}$. This answers (a.) and also shows that he eats chicken and vegetable soup each with probability $\frac{1}{6}$.

To find the long-run probabilities, we must find the steady-state vector $\mathbf{s}$. Theorem 2.9.2 applies because P is regular (P^2 has positive entries), so $\mathbf{s}$ satisfies $P\mathbf{s} = \mathbf{s}$. That is, $(I - P)\mathbf{s} = \mathbf{0}$ where

$$I - P = \frac{1}{6}\begin{bmatrix} 6 & -4 & -4 \\ -3 & 6 & -2 \\ -3 & -2 & 6 \end{bmatrix}$$

The solution is $\mathbf{s} = \begin{bmatrix} 4t \\ 3t \\ 3t \end{bmatrix}$, where t is a parameter, and we use $\mathbf{s} = \begin{bmatrix} 0.4 \\ 0.3 \\ 0.3 \end{bmatrix}$ because the entries of $\mathbf{s}$ must sum to 1. Hence, in the long run, he eats beef soup 40% of the time and eats chicken soup and vegetable soup each 30% of the time.

---

1. The interested reader can find an elementary proof in J. Kemeny, H. Mirkil, J. Snell, and G. Thompson, *Finite Mathematical Structures* (Englewood Cliffs, N.J.: Prentice-Hall, 1958).

This page titled 7.2: An Application to Markov Chains is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by W. Keith Nicholson via source content that was edited to the style and standards of the LibreTexts platform.

- 2.9: An Application to Markov Chains by W. Keith Nicholson is licensed CC BY-NC-SA 4.0. Original source: https://lyryx.com/linear-algebra-applications.

7.2.1: Absorbing Markov Chains

Learning Objectives

In this section, you will learn to:

1. Identify absorbing states and absorbing Markov chains
2. Solve and interpret absorbing Markov chains.

In this section, we will study a type of Markov chain in which when a certain state is reached, it is impossible to leave that state. Such states are called **absorbing states**, and a Markov Chain that has at least one such state is called an **Absorbing Markov chain.**

Suppose you have the following transition matrix.

$$\begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \end{array} \begin{array}{c} \begin{array}{ccc} S_1 & S_2 & S_3 \end{array} \\ \begin{bmatrix} .1 & .3 & .6 \\ 0 & 1 & 0 \\ .3 & .2 & .5 \end{bmatrix} \end{array}$$

The state S_2 is an absorbing state, because the probability of moving from state S_2 to state S_2 is 1. Which is another way of saying that if you are in state S_2, you will remain in state S_2.

In fact, this is the way to identify an absorbing state. If the probability in row i and column i, p_{ii}, is 1, then state S_i is an absorbing state.

Example 7.2.1.1

Consider transition matrices A, B, C for Markov chains shown below. Which of the following Markov chains have an absorbing state?

$$A = \begin{bmatrix} .3 & .7 & 0 \\ 0 & 1 & 0 \\ .2 & .3 & .5 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad C = \begin{bmatrix} .1 & .3 & .4 & .2 \\ 0 & .2 & .1 & .7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Solution

$$\begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \end{array} \begin{array}{c} \begin{array}{ccc} S_1 & S_2 & S_3 \end{array} \\ \begin{bmatrix} .3 & .7 & 0 \\ 0 & 1 & 0 \\ .2 & .3 & .5 \end{bmatrix} \end{array}$$

A has S_2 as an absorbing state.

If we are in state S_2, we can not leave it. From state S_2, we can not transition to state S_1 or S_3; the probabilities are 0. The probability of transition from state S_2 to state S_2 is 1.

$$B = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \end{array} \begin{array}{c} \begin{array}{ccc} S_1 & S_2 & S_3 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$ does not have any absorbing states.

From state S_1, we always transition to state S_2. From state S_2 we always transition to state S_3. From state S_3 we always transition to state S_1. In this matrix, it is never possible to stay in the same state during a transition.

$$C = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \end{array} \begin{array}{c} \begin{array}{cccc} S_1 & S_2 & S_3 & S_4 \end{array} \\ \begin{bmatrix} .1 & .3 & .4 & .2 \\ 0 & .2 & .1 & .7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$ has two absorbing states, S_3 and S_4.

From state S_3, you can only remain in state S_3, and never transition to any other states. Similarly from state S_4, you can only remain in state S_4, and never transition to any other states.

We summarize how to identify absorbing states.

Summary

A state S is an absorbing state in a Markov chain in the transition matrix if

- The row for state S has one 1 and all other entries are 0

AND

- The entry that is 1 is on the main diagonal (row = column for that entry), indicating that we can never leave that state once it is entered.

Next we define an absorbing Markov Chain

Absorbing Markov Chain

A Markov chain is an absorbing Markov Chain if

- It has at least one absorbing state

AND

- From any non-absorbing state in the Markov chain, it is possible to eventually move to some absorbing state (in one or more transitions).

Example 7.2.1.2

Consider transition matrices C and D for Markov chains shown below. Which of the following Markov chains is an absorbing Markov Chain?

$$C = \begin{bmatrix} .1 & .3 & .4 & .2 \\ 0 & .2 & .1 & .7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ .2 & .2 & .2 & .2 & .2 \\ 0 & 0 & 0 & .3 & .7 \\ 0 & 0 & 0 & .6 & .4 \end{bmatrix}$$

Solution

C is an absorbing Markov Chain but D is not an absorbing Markov chain.

Matrix C has two absorbing states, S_3 and S_4, and it is possible to get to state S_3 and S_4 from S_1 and S_2.

Matrix D is not an absorbing Markov chain. It has two absorbing states, S_1 and S_2, but it is never possible to get to either of those absorbing states from either S_4 or S_5. If you are in state S_4 or S_5, you always remain transitioning between states S_4 or S_{5m} and can never get absorbed into either state S_1 or S_2

In the remainder of this section, we'll examine absorbing Markov chains with two classic problems: the random drunkard's walk problem and the gambler's ruin problem. And finally we'll conclude with an absorbing Markov model applied to a real world situation.

Drunkard's Random Walk

In this example we briefly examine the basic ideas concerning absorbing Markov chains.

Example 7.2.1.3

A man walks along a three block portion of Main St. His house is at one end of the three block section. A bar is at the other end of the three block section. Each time he reaches a corner he randomly either goes forward one block or turns around and goes back one block. If he reaches home or the bar, he stays there. The four states are Home (H), Corner 1(C_1), Corner 2 (C_2) and Bar (B).

Write the transition matrix and identify the absorbing states. Find the probabilities of ending up in each absorbing state depending on the initial state.

Solution

The transition matrix is written below.

$$T = \begin{array}{c} \\ H \\ C_1 \\ C_2 \\ B \end{array} \begin{array}{c} \begin{array}{cccc} H & C_1 & C_2 & B \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ .5 & 0 & .5 & 0 \\ 0 & .5 & 0 & .5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Home and the Bar are absorbing states. If the man arrives home, he does not leave. If the man arrives at the bar, he does not leave. Since it is possible to reach home or the bar from each of the other two corners on his walk, this is an absorbing Markov chain.

We can raise the transition matrix T to a high power, n. One we find a power T^n that remains stable, it will tell us the probability of ending up in each absorbing state depending on the initial state.

$$T^{90} = \begin{array}{c} \\ H \\ C_1 \\ C_2 \\ B \end{array} \begin{array}{c} \begin{array}{cccc} H & C_1 & C_2 & B \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2/3 & 0 & 0 & 1/3 \\ 1/3 & 0 & 0 & 2/3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \quad \text{and} \quad T^{91} = \begin{array}{c} \\ H \\ C_1 \\ C_2 \\ B \end{array} \begin{array}{c} \begin{array}{cccc} H & C_1 & C_2 & B \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2/3 & 0 & 0 & 1/3 \\ 1/3 & 0 & 0 & 2/3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

$T^{91} = T^{90}$; the matrix does not change as we continue to examine higher powers. We see that in the long-run, the Markov chain must end up in an absorbing state. In the long run, the man must eventually end up at either his home or the bar.

The second row tells us that if the man is at corner C_1, then there is a 2/3 chance he will end up at home and a 1/3 chance he will end up at the bar.

The third row tells us that if the man is at corner C_2, then there is a 1/3 chance he will end up at home and a 2/3 chance he will end up at the bar.

Once he reaches home or the bar, he never leaves that absorbing state.

Note that while the matrix T^n for sufficiently large n has become stable and is not changing, it does not represent an equilibrium matrix. The rows are not all identical, as we found in the regular Markov chains that reached an equilibrium.

We can write a smaller "solution matrix" by retaining only rows that relate to the non-absorbing states and retaining only the columns that relate to the absorbing states.

Then the solution matrix will have rows C_1 and C_2, and columns H and B. The solution matrix is

$$\text{Solution Matrix:} \quad \begin{array}{c} \\ C_1 \\ C_2 \end{array} \begin{array}{c} \begin{array}{cc} H & B \end{array} \\ \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix} \end{array}$$

The first row of the solution matrix shows that if the man is at corner C_1, then there is a 2/3 chance he will end up at home and a 1/3 chance he will end up at the bar.

The second row of the solution matrix shows that if the man is at corner C_2, then there is a 1/3 chance he will end up at home and a 2/3 chance he will end up at the bar.

The solution matrix does not show that eventually there is 0 probability of ending up in C1 or C2, or that if you start in an absorbing state H or B, you stay there. The smaller solution matrix assumes that we understand these outcomes and does not include that information.

The next example is another classic example of an absorbing Markov chain. In the next example we examine more of the mathematical details behind the concept of the solution matrix.

Gamber's Ruin Problem

Example 7.2.1.4

A gambler has $3,000, and she decides to gamble $1,000 at a time at a Black Jack table in a casino in Las Vegas. She has told herself that she will continue playing until she goes broke or has $5,000. Her probability of winning at Black Jack is .40. Write the transition matrix, identify the absorbing states, find the solution matrix, and determine the probability that the gambler will be financially ruined at a stage when she has $2,000.

Solution

The transition matrix is written below. Clearly the state 0 and state 5K are the absorbing states. This makes sense because as soon as the gambler reaches 0, she is financially ruined and the game is over. Similarly, if the gambler reaches $5,000, she has promised herself to quit and, again, the game is over. The reader should note that $p_{00} = 1$, and $p_{55} = 1$.

Further observe that since the gambler bets only $1,000 at a time, she can raise or lower her money only by $1,000 at a time. In other words, if she has $2,000 now, after the next bet she can have $3,000 with a probability of .40 and $1,000 with a probability of .60.

	0	1K	2K	3K	4K	5K
0	1	0	0	0	0	0
1K	.60	0	.40	0	0	0
2K	0	.60	0	.40	0	0
3K	0	0	.60	0	.40	0
4K	0	0	0	.60	0	.40
5K	0	0	0	0	0	1

To determine the long term trend, we raise the matrix to higher powers until all the non-absorbing states are absorbed. This is the called the **solution matrix.**

	0	1K	2K	3K	4K	5K
0	1	0	0	0	0	0
1K	195/211	0	0	0	0	16/211
2K	171/211	0	0	0	0	40/211
3K	135/211	0	0	0	0	76/211
4K	81/211	0	0	0	0	130/211
5K	0	0	0	0	0	1

The solution matrix is often written in the following form, where the non-absorbing states are written as rows on the side, and the absorbing states as columns on the top.

	0	5K
1K	195/211	16/211
2K	171/211	40/211
3K	135/211	76/211
4K	81/211	130/211

The table lists the probabilities of getting absorbed in state 0 or state 5K starting from any of the four non-absorbing states. For example, if at any instance the gambler has $3,000, then her probability of financial ruin is 135/211 and her probability reaching 5K is 76/211.

Example 7.2.1.5

Solve the Gambler's Ruin Problem of Example 7.2.1.4 without raising the matrix to higher powers, and determine the number of bets the gambler makes before the game is over.

Solution

In solving absorbing states, it is often convenient to rearrange the matrix so that the rows and columns corresponding to the absorbing states are listed first. This is called the **Canonical form.** The transition matrix of Example 1 in the canonical form is listed below.

	0	5K	1K	2K	3K	4K
0	1	0	0	0	0	0
5K	0	1	0	0	0	0
1K	.60	0	0	.40	0	0
2K	0	0	.60	0	.40	0
3K	0	0	0	.60	0	.40
4K	0	.40	0	0	.60	0

The canonical form divides the transition matrix into four sub-matrices as listed below.

	Absorbing	Non-absorbing
Absorbing states	I_n	O
Non-absorbing states	A	B

The matrix $F = (I_n - B)^{-1}$ is called the fundamental matrix for the absorbing Markov chain, where In is an identity matrix of the same size as B. The i, j-th entry of this matrix tells us the average number of times the process is in the non-absorbing state j before absorption if it started in the non-absorbing state i.

The matrix $F = (I_n - B)^{-1}$ for our problem is listed below.

$$F = \begin{array}{c} \\ 1K \\ 2K \\ 3K \\ 4K \end{array} \begin{array}{c} \begin{array}{cccc} 1K & 2K & 3K & 4K \end{array} \\ \begin{bmatrix} 1.54 & .90 & .47 & .19 \\ 1.35 & 2.25 & 1.18 & .47 \\ 1.07 & 1.78 & 2.25 & .90 \\ .64 & 1.07 & 1.35 & 1.54 \end{bmatrix} \end{array}$$

You can use your calculator, or a computer, to calculate matrix F.

The Fundamental matrix F helps us determine the average number of games played before absorption.

According to the matrix, the entry 1.78 in the row 3, column 2 position says that the gambler will play the game 1.78 times before she goes from \$3K to \$2K. The entry 2.25 in row 3, column 3 says that if the gambler now has \$3K, she will have \$3K on the average 2.25 times before the game is over.

We now address the question of how many bets will she have to make before she is absorbed, if the gambler begins with \$3K?

If we add the number of games the gambler plays in each non-absorbing state, we get the average number of games before absorption from that state. Therefore, if the gambler starts with \$3K, the average number of Black Jack games she will play before absorption is

$$1.07 + 1.78 + 2.25 + 0.90 = 6.0$$

That is, we expect the gambler will either have \$5,000 or nothing on the 7th bet.

Lastly, we find the solution matrix without raising the transition matrix to higher powers. The matrix FA gives us the solution matrix.

$$\mathrm{FA} = \begin{bmatrix} 1.54 & .90 & .47 & .19 \\ 1.35 & 2.25 & 1.18 & .47 \\ 1.07 & 1.78 & 2.25 & .90 \\ .64 & 1.07 & 1.35 & 1.54 \end{bmatrix} \begin{bmatrix} .6 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & .4 \end{bmatrix} = \begin{bmatrix} .92 & .08 \\ .81 & .19 \\ .64 & .36 \\ .38 & .62 \end{bmatrix}$$

which is the same as the following matrix we obtained by raising the transition matrix to higher powers.

$$\begin{array}{c} \\ 1K \\ 2K \\ 3K \\ 4K \end{array} \begin{array}{c} \begin{array}{cc} 0 & 5K \end{array} \\ \begin{bmatrix} 195/211 & 16/211 \\ 171/211 & 40/211 \\ 135/211 & 76/211 \\ 81/211 & 130/211 \end{bmatrix} \end{array}$$

Example 7.2.1.6

At a professional school, students need to take and pass an English writing/speech class in order to get their professional degree. Students must take the class during the first quarter that they enroll. If they do not pass the class they take it again in the second semester. If they fail twice, they are not permitted to retake it again, and so would be unable to earn their degree.

Students can be in one of 4 states: passed the class (P), enrolled in the class for the first time (C), retaking the class (R) or failed twice and can not retake (F). Experience shows 70% of students taking the class for the first time pass and 80% of students taking the class for the second time pass.

Write the transition matrix and identify the absorbing states. Find the probability of being absorbed eventually in each of the absorbing states.

Solution

The absorbing states are P (pass) and F (fail repeatedly and can not retake). The transition matrix T is shown below.

$$T=\begin{array}{c} \\ P \\ C \\ R \\ F \end{array}\begin{array}{c} \begin{array}{cccc} P & C & R & F \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ .7 & 0 & .3 & 0 \\ .8 & 0 & 0 & .2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

If we raise the transition matrix T to a high power, we find that it remains stable and gives us the long-term probabilities of ending up in each of the absorbing states.

$$T^{30}=\begin{array}{c} \\ P \\ C \\ R \\ F \end{array}\begin{array}{c} \begin{array}{cccc} P & C & R & F \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ .94 & 0 & 0 & .06 \\ .8 & 0 & 0 & .2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Of students currently taking the class for the first time, 94% will eventually pass. 6% will eventually fail twice and be unable to earn their degree.

Of students currently taking the class for the second time time, 80% will eventually pass. 20% will eventually fail twice and be unable to earn their degree.

The solution matrix contains the same information in abbreviated form

$$\text{Solution Matrix}=\begin{array}{c} \\ C \\ R \end{array}\begin{array}{c} \begin{array}{cc} P & F \end{array} \\ \begin{bmatrix} .94 & .06 \\ .80 & .20 \end{bmatrix} \end{array}$$

Note that in this particular problem, we don't need to raise T to a "very high" power. If we find T^2, we see that it is actually equal to T^n for higher powers n. T^n becomes stable after two transitions; this makes sense in this problem because after taking the class twice, the student must have passed or is not permitted to retake it any longer. Therefore the probabilities should not change any more after two transitions; by the end of two transitions, every student has reached an absorbing state.

Absorbing Markov Chains

1. A Markov chain is an absorbing Markov chain if it has at least one absorbing state. A state i is an absorbing state if once the system reaches state i, it stays in that state; that is, $p_{ii} = 1$.
2. If a transition matrix T for an absorbing Markov chain is raised to higher powers, it reaches an absorbing state called the solution matrix and stays there. The i, j-th entry of this matrix gives the probability of absorption in state j while starting in state i.
3. Alternately, the solution matrix can be found in the following manner.
   a. Express the transition matrix in the canonical form as below.

$$\mathrm{T} = \begin{bmatrix} \mathbf{I_n} & \mathbf{0} \\ \mathbf{A} & \mathbf{B} \end{bmatrix}$$

   where I_n is an identity matrix, and **0** is a matrix of all zeros.
   b. The fundamental matrix $F = (I - B)^{-1}$. The fundamental matrix helps us find the number of games played before absorption.
   c. FA is the solution matrix, whose i, j-th entry gives the probability of absorption in state j while starting in state i.
4. The sum of the entries of a row of the fundamental matrix gives us the expected number of steps before absorption for the non-absorbing state associated with that row.

This page titled 7.2.1: Absorbing Markov Chains is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rupinder Sekhon and Roberta Bloom via source content that was edited to the style and standards of the LibreTexts platform.

- 10.4: Absorbing Markov Chains by Rupinder Sekhon and Roberta Bloom is licensed CC BY 4.0. Original source: https://www.deanza.edu/faculty/bloomroberta/math11/afm3files.html.html.

7.2.1.1: Absorbing Markov Chains (Exercises)

SECTION 10.4 PROBLEM SET: ABSORBING MARKOV CHAINS

1. Given the following absorbing Markov chain.

$$T = \begin{array}{c} \\ S1 \\ S2 \\ S3 \\ S4 \end{array} \begin{array}{c} \begin{array}{cccc} S1 & S2 & S3 & S4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ .1 & .4 & .2 & .3 \\ 0 & 0 & 1 & 0 \\ .4 & 0 & .2 & .4 \end{bmatrix} \end{array}$$

a. Identify the absorbing states.	b. Write the solution matrix.
c. Starting from state 4, what is the probability of eventual absorption in state 1?	d. Starting from state 2, what is the probability of eventual absorption in state 3?

2. Two tennis players, Andre and Vijay each with two dollars in their pocket, decide to bet each other $1, for every game they play. They continue playing until one of them is broke.

a. Write the transition matrix for Andre.	b. Identify the absorbing states.
c. Write the solution matrix.	d. At a given stage if Andre has $1, what is the chance that he will eventually lose it all?

3. Repeat the previous problem, if the chance of winning for Andre is .4 and for Vijay .6.

a. Write the transition matrix for Andre.	b. Identify the absorbing states.
c. If Andre has $3, what is the probability that he will eventually be ruined?	d. If Vijay has $1, what is the probability that he will eventually triumph?

4. Repeat problem 2, if initially Andre has $3 and Vijay has $2.

a. Write the transition matrix.	b. Identify the absorbing states.
c. Write the solution matrix	d. If Andre has $4, what is the probability that he will eventually be ruined?

5. The non-tenured professors at a community college are regularly evaluated. After an evaluation they are classified as good, bad, or improvable. The "improvable" are given a set of recommendations and are re-evaluated the following semester. At the next evaluation, 60% of the improvable turn out to be good, 20% bad, and 20% improvable. These percentages never change and the process continues.

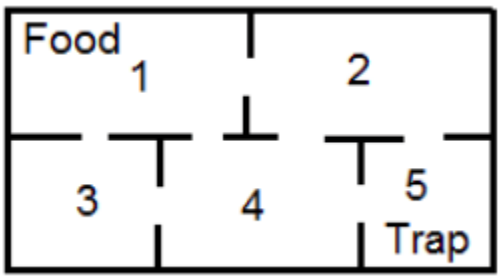
a. Write the transition matrix.	b. Identify the absorbing states.
c. Write the solution matrix	d. What is the probability that a professor who is improvable will eventually become good?

Questions 6 - 11 refer to the following:
In a professional certification program students take classes and then participate in an internships.
There are 4 states: taking classes (C), internship (I), drop out (D), and graduate (G).
If a student drops out they are never readmitted to the program.

Of those students currently taking classes, 70% have an internship the next year, 20% are still taking classes the next year, and 10% have dropped out by the next year. Of the students who are currently doing an internship, 65% graduate by the next year; 20% drop out by the next year, and 15% are still completing their internship the next year.

6. Write the transition matrix and indicate which are the absorbing states.	7. If a student is taking classes now: a. find the probability that the student will graduate in 2 years b. find the probability that the student will be in the internship in 2 years. c. find the probability that the student will have dropped out by 2 years from now.
8. Find the probability that a student currently doing an internship will eventually drop out.	9. Find the probability that a student taking classes now will eventually graduate.
10. If 40% of students are currently taking classes and 60% of current students are doing internships, what is the eventual long term distribution of students for graduating versus dropping out?	11. If 70% of students are currently taking classes and 30% of current students are doing internships, what is the eventual long term distribution of students for graduating versus dropping out?

12. A mouse is placed in the maze shown below, and it moves from room to room randomly. From any room, the mouse will choose a door to the next room with equal probabilities. Once the mouse reaches room 1, it finds food and never leaves that room. And when it reaches room 5, it is trapped and cannot leave that room. What is the probability the mouse will end up in room 5 if it was initially placed in room 3?



13. In problem 12, what is the probability the mouse will end up in room 1 if initially placed in room 2?

This page titled 7.2.1.1: Absorbing Markov Chains (Exercises) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rupinder Sekhon and Roberta Bloom via source content that was edited to the style and standards of the LibreTexts platform.

- 10.4.1: Absorbing Markov Chains (Exercises) by Rupinder Sekhon and Roberta Bloom is licensed CC BY 4.0. Original source: https://www.deanza.edu/faculty/bloomroberta/math11/afm3files.html.html.

7.2.2: Introduction to Markov Chains

Learning Objectives

In this chapter, you will learn to:

1. Write transition matrices for Markov Chain problems.
2. Use the transition matrix and the initial state vector to find the state vector that gives the distribution after a specified number of transitions.

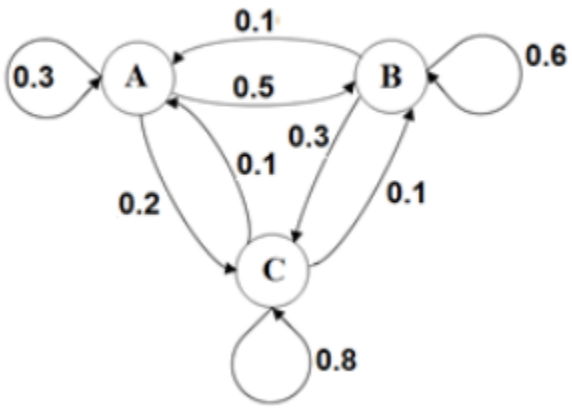
We will now study stochastic processes, experiments in which the outcomes of events depend on the previous outcomes; stochastic processes involve random outcomes that can be described by probabilities. Such a process or experiment is called a **Markov Chain** or **Markov process**. The process was first studied by a Russian mathematician named Andrei A. Markov in the early 1900s.

About 600 cities worldwide have bike share programs. Typically a person pays a fee to join a the program and can borrow a bicycle from any bike share station and then can return it to the same or another system. Each day, the distribution of bikes at the stations changes, as the bikes get returned to different stations from where they are borrowed.

For simplicity, let's consider a very simple bike share program with only 3 stations: A, B, C. Suppose that all bicycles must be returned to the station at the end of the day, so that each day there is a time, let's say midnight, that all bikes are at some station, and we can examine all the stations at this time of day, every day. We want to model the movement of bikes from midnight of a given day to midnight of the next day. We find that over a 1 day period,

- of the bikes borrowed from station A, 30% are returned to station A, 50% end up at station B, and 20% end up at station C.
- of the bikes borrowed from station B, 10% end up at station A, 60% have been returned to station B, and 30% end up at station C
- of the bikes borrowed from station C, 10% end up at station A, 10% end up at station B, and 80% are returned to station C.

We can draw an arrow diagram to show this. The arrows indicate the station where the bicycle was started, called its initial state, and the stations at which it might be located one day later, called the terminal states. The numbers on the arrows show the probability for being in each of the indicated terminal states.



Because our bike share example is simple and has only 3 stations, the arrow diagram, also called a directed graph, helps us visualize the information. But if we had an example with 10, or 20, or more bike share stations, the diagram would become so complicated that it would be difficult to understand the information in the diagram.

We can use a **transition matrix** to organize the information,

Each row in the matrix represents an initial state. Each column represents a terminal state.
We will assign the rows in order to stations A, B, C, and the columns in the same order to stations A, B, C. Therefore the matrix must be a square matrix, with the same number of rows as columns. The entry in row 2 column 3, for example, would show the probability that a bike that is initially at station B will be at station C one day later: that entry is 0.30, which is the probability in the diagram for the arrow that points from B to C. We use this the letter T for transition matrix.

$$T = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \end{array}$$

Looking at the first row that represents bikes initially at station A, we see that 30% of the bikes borrowed from station A are returned to station A, 50% end up at station B, and 20% end up at station C, after one day.

We note some properties of the transition matrix:

- t_{ij} represents the entry in row i column j
- t_{ij} = the probability of moving from state represented by row i to the state represented by row j in a single transition

- t_{ij} is a conditional probability which we can write as:
  t_{ij} = P(next state is the state in column j | current state is the state in row i)
- Each row adds to 1
- All entries are between 0 and 1, inclusive because they are probabilities.
- The transition matrix represents change over one transition period; in this example one transition is a fixed unit of time of one day.

Example 7.2.2.1

A city is served by two cable TV companies, BestTV and CableCast.

- Due to their aggressive sales tactics, each year 40% of BestTV customers switch to CableCast; the other 60% of BestTV customers stay with BestTV.
- On the other hand, 30% of the CableCast customers switch to Best TV.

The two states in this example are BestTV and CableCast. Express the information above as a transition matrix which displays the probabilities of going from one state into another state.

Solution

The transition matrix is:

		Next year	
		BestTV	CableCast
This year	BestTV	.60	.40
	CableCast	.30	.70

As previously noted, the reader should observe that a transition matrix is always a square matrix because all possible states must have both rows and columns. All entries in a transition matrix are non-negative as they represent probabilities. And, since all possible outcomes are considered in the Markov process, the sum of the row entries is always 1.

With a larger transition matrix, the ideas in Example 7.2.2.1 could be expanded to represent a market with more than 2 cable TV companies. The concepts of brand loyalty and switching between brands demonstrated in the cable TV example apply to many types of products, such as cell phone carriers, brands of regular purchases such as food or laundry detergent, brands major purchases such as cars or appliances, airlines that travelers choose when booking flights, or hotels chains that travelers choose to stay in.

The transition matrix shows the probabilities for transitions between states at two consecutive times. We need a way to represent the distribution among the states at a particular point in time. To do this we use a row matrix called a **state vector**. The state vector is a row matrix that has only one row; it has one column for each state. The entries show the distribution by state at a given point in time. All entries are between 0 and 1 inclusive, and the sum of the entries is 1.

For the bike share example with 3 bike share stations, the state vector is a 1×3 matrix with 1 row and 3 columns. Suppose that when we start observing our bike share program, 30% of the bikes are at station A, 45% of the bikes are at station B, and 25% are at station C. The initial state vector is

$$\begin{matrix} & \text{A} & \text{B} & \text{C} \end{matrix}$$
$$\text{V}_0 = \begin{bmatrix} 0.30 & 0.45 & 0.25 \end{bmatrix}$$

The subscript 0 indicates that this is the initial distribution, before any transitions occur.

If we want to determine the distribution after one transition, we'll need to find a new state vector that we'll call V_1. The subscript 1 indicates this is the distribution after 1 transition has occurred.

We find V_1 by multiplying V_0 by the transition matrix T, as follows:

$$\begin{aligned} &\text{V}_1 = \text{V}_0\text{T} \\ &= \begin{bmatrix} 0.30 & 0.45 & 0.25 \end{bmatrix} \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \\ &= [.30(.3) + .45(.1) + .25(.1) \quad .30(.5) + .45(.6) + .25(.1) \quad .30(.2) + .45(.3) + .25(.8)] \\ &= \begin{bmatrix} .16 & .445 & .395 \end{bmatrix} \end{aligned}$$

After 1 day (1 transition), 16 % of the bikes are at station A, 44.5 % are at station B and 39.5% are at station C.

We showed the step by step work for the matrix multiplication above. In the future we'll generally use technology, such as the matrix capabilities of our calculator, to perform any necessary matrix multiplications, rather than showing the step by step work.

Suppose now that we want to know the distribution of bicycles at the stations after two days. We need to find V_2, the state vector after two transitions. To find V_2 , we multiply the state vector after one transition V_1 by the transition matrix T.

$$\text{V}_2 = \text{V}_1\text{T} = \begin{bmatrix} .16 & .445 & .395 \end{bmatrix} \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} = \begin{bmatrix} .132 & .3865 & .4815 \end{bmatrix}$$

We note that $\text{V}_1 = \text{V}_0\text{T}, \text{ so } \text{V}_2 = \text{V}_1\text{T} = (\text{V}_0\text{T})\,\text{T} = \text{V}_0\text{T}^2$

This gives an equivalent method to calculate the distribution of bicycles on day 2:

$$\begin{aligned} &\text{V}_2 = \text{V}_0\text{T}^2 = \begin{bmatrix} 0.30 & 0.45 & 0.25 \end{bmatrix} \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}^2 \\ &= \begin{bmatrix} 0.30 & 0.45 & 0.25 \end{bmatrix} \begin{bmatrix} 0.16 & 0.47 & 0.37 \\ 0.12 & 0.44 & 0.44 \\ 0.12 & 0.19 & 0.69 \end{bmatrix} \\ &\text{V}_2 = \text{V}_0\text{T}^2 = \begin{bmatrix} .132 & .3865 & .4815 \end{bmatrix} \end{aligned}$$

After 2 days (2 transitions), 13.2 % of the bikes are at station A, 38.65 % are at station B and 48.15% are at station C.

We need to examine the following: What is the meaning of the entries in the matrix T^2?

$$T^2 = TT = \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.16 & 0.47 & 0.37 \\ 0.12 & 0.44 & 0.44 \\ 0.12 & 0.19 & 0.69 \end{bmatrix}$$

The entries in T^2 tell us the probability of a bike being at a particular station after two transitions, given its initial station.

- Entry t_{13} in row 1 column 3 tells us that a bike that is initially borrowed from station A has a probability of 0.37 of being in station C after two transitions.
- Entry t_{32} in row 3 column 2 tells us that a bike that is initially borrowed from station C has a probability of 0. 19 of being in station B after two transitions.

Similarly, if we raise transition matrix T to the nth power, the entries in T^n tells us the probability of a bike being at a particular station after n transitions, given its initial station.

And if we multiply the initial state vector V_0 by T^n, the resulting row matrix $Vn=V_0T^n$ is the distribution of bicycles after n transitions.

Example 7.2.2.2

Refer to Example 7.2.2.1 for the transition matrix for market shares for subscribers to two cable TV companies.

a. Suppose that today 1/4 of customers subscribe to BestTV and 3/4 of customers subscribe to CableCast. After 1 year, what percent subscribe to each company?
b. Suppose instead that today of 80% of customers subscribe to BestTV and 20% subscribe to CableCast. After 1 year, what percent subscribe to each company?

Solution

a.The initial distribution given by the initial state vector is a 1×2 matrix

$$V_0 = \begin{bmatrix} 1/4 & 3/4 \end{bmatrix} = \begin{bmatrix} .25 & .75 \end{bmatrix}$$

and the transition matrix is

$$T = \begin{bmatrix} .60 & .40 \\ .30 & .70 \end{bmatrix}$$

After 1 year, the distribution of customers is

$$V_1 = V_0T = \begin{bmatrix} .25 & .75 \end{bmatrix} \begin{bmatrix} .60 & .40 \\ .30 & .70 \end{bmatrix} = \begin{bmatrix} .375 & .625 \end{bmatrix}$$

After 1 year, 37.5% of customers subscribe to BestTV and 62.5% to CableCast.

b. The initial distribution given by the initial state vector $V_0 = \begin{bmatrix} .8 & .2 \end{bmatrix}$. Then

$$V_1 = V_0T = \begin{bmatrix} .8 & .2 \end{bmatrix} \begin{bmatrix} .60 & .40 \\ .30 & .70 \end{bmatrix} = \begin{bmatrix} .54 & .46 \end{bmatrix}$$

In this case, after 1 year, 54% of customers subscribe to BestTV and 46% to CableCast.

Note that the distribution after one transition depends on the initial distribution; the distributions in parts (a) and (b) are different because of the different initial state vectors.

Example 7.2.2.3

Professor Symons either walks to school, or he rides his bicycle. If he walks to school one day, then the next day, he will walk or cycle with equal probability. But if he bicycles one day, then the probability that he will walk the next day is 1/4. Express this information in a transition matrix.

Solution

We obtain the following transition matrix by properly placing the row and column entries. Note that if, for example, Professor Symons bicycles one day, then the probability that he will walk the next day is 1/4, and therefore, the probability that he will bicycle the next day is 3/4.

		Next Day	
		Walk	Bicycle
Initial Day	Walk	1/2	1/2
	Bicycle	1/4	3/4

Example 7.2.2.4

In Example 7.2.2.3, if it is assumed that the initial day is Monday, write a matrix that gives probabilities of a transition from Monday to Wednesday.

Solution

If today is Monday, then Wednesday is two days from now, representing two transitions. We need to find the square, T^2, of the original transition matrix T, using matrix multiplication.

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

$$\begin{aligned} T^2 = T \times T &= \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix} \begin{bmatrix} 1/2 & \\ 1/4 & 1/2 \end{bmatrix} \\ &= \begin{bmatrix} 1/4+1/8 & 1/4+3/8 \\ 1/8+3/16 & 1/8+9/16 \end{bmatrix} \\ &= \begin{bmatrix} 3/8 & 5/8 \\ 5/16 & 11/16 \end{bmatrix} \end{aligned} \tag{7.2.2.1}$$

Recall that we do not obtain T^2 by squaring each entry in matrix T, but obtain T^2 by multiplying matrix T by itself using matrix multiplication.

We represent the results in the following matrix.

		Wednesday	
		Walk	Bicycle
Monday	Walk	3/8	5/8
	Bicycle	5/16	11/16

We interpret the probabilities from the matrix T^2 as follows:

- P(Walked Wednesday | Walked Monday) = 3/8.
- P(Bicycled Wednesday | Walked Monday) = 5/8.
- P(Walked Wednesday | Bicycled Monday) = 5/16.
- P(Bicycled Wednesday | Bicycled Monday) = 11/16.

Example 7.2.2.5

The transition matrix for Example 7.2.2.3 is given below.

		Tuesday	
		Walk	Bicycle
Monday	Walk	1/2	1/2
	Bicycle	1/4	3/4

Write the transition matrix from

a. Monday to Thursday
b. Monday to Friday.

Solution

a.) In writing a transition matrix from Monday to Thursday, we are moving from one state to another in three steps. That is, we need to compute T^3.

$$T^3 = \begin{bmatrix} 11/32 & 21/32 \\ 21/64 & 43/64 \end{bmatrix}$$

b) To find the transition matrix from Monday to Friday, we are moving from one state to another in 4 steps. Therefore, we compute T^4.

$$T^4 = \begin{bmatrix} 43/128 & 85/128 \\ 85/256 & 171/256 \end{bmatrix}$$

It is important that the student is able to interpret the above matrix correctly. For example, the entry 85/128, states that if Professor Symons walked to school on Monday, then there is 85/128 probability that he will bicycle to school on Friday.

There are certain Markov chains that tend to stabilize in the long run. We will examine these more deeply later in this chapter. The transition matrix we have used in the above example is just such a Markov chain. The next example deals with the long term trend or steady-state situation for that matrix.

Example 7.2.2.6

Suppose Professor Symons continues to walk and bicycle according to the transition matrix given in Example 7.2.2.3. In the long run, how often will he walk to school, and how often will he bicycle?

Solution

If we examine higher powers of the transition matrix T, we will find that it stabilizes.

$$T^5 = \begin{bmatrix} .333984 & 666015 \\ .333007 & .666992 \end{bmatrix} \quad T^{10} = \begin{bmatrix} .33333397 & .66666603 \\ .333333301 & .66666698 \end{bmatrix}$$

$$\text{And} \quad T^{20} = \begin{bmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{bmatrix} \quad \text{and } T^{n} = \begin{bmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{bmatrix} \text{ for } n > 20$$

The matrix shows that in the long run, Professor Symons will walk to school 1/3 of the time and bicycle 2/3 of the time.

When this happens, we say that the system is in steady-state or state of equilibrium. In this situation, all row vectors are equal. If the original matrix is an n by n matrix, we get n row vectors that are all the same. We call this vector a **fixed probability vector** or the **equilibrium vector** E. In the above problem, the fixed probability vector E is [1/3 2/3]. Furthermore, if the equilibrium vector E is multiplied by the original matrix T, the result is the equilibrium vector E. That is,

ET = E , or $\begin{bmatrix} 1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/3 \end{bmatrix}$

This page titled 7.2.2: Introduction to Markov Chains is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rupinder Sekhon and Roberta Bloom via source content that was edited to the style and standards of the LibreTexts platform.

- 10.1: Introduction to Markov Chains by Rupinder Sekhon and Roberta Bloom is licensed CC BY 4.0. Original source: https://www.deanza.edu/faculty/bloomroberta/math11/afm3files.html.html.

7.2.2.1: Introduction to Markov Chains (Exercises)

SECTION 10.1 PROBLEM SET: INTRODUCTION TO MARKOV CHAINS

1. Is the matrix given below a transition matrix for a Markov chain? Explain.

a. $\begin{bmatrix} .2 & .3 & .5 \\ .3 & -.2 & .9 \\ .3 & .3 & .5 \end{bmatrix}$

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

2. A survey of American car buyers indicates that if a person buys a Ford, there is a 60% chance that their next purchase will be a Ford, while owners of a GM will buy a GM again with a probability of .80. The buying habits of these consumers are represented in the transition matrix below.

		Next Purchase: Ford	Next Purchase: GM
Present Purchase	Ford	.60	.40
	GM	.20	.80

Find the following probabilities:

a. The probability that a present owner of a Ford will buy a GM as his next car.

b. The probability that a present owner of a GM will buy a GM as his next car.

c. The probability that a present owner of a Ford will buy a GM as his third car.

d. The probability that a present owner of a GM will buy a GM as his fourth car.

3. Professor Hay has breakfast at Hogee's every morning. He either orders an Egg Scramble, or a Tofu Scramble. He never orders Eggs on two consecutive days, but if he does order Tofu one day, then the next day he can order Tofu or Eggs with equal probability.

a. Write a transition matrix for this problem.

b. If Professor Hay has Tofu on Monday, what is the probability he will have Tofu on Tuesday?

c. If Professor Hay has Eggs on Monday, find the probability he will have Tofu on Wednesday.

d. If Professor Hay has Eggs on Monday, what is the probability he will have Tofu on Thursday?

4. A professional tennis player always hits cross-court or down the line. In order to give himself a tactical edge, he never hits down the line two consecutive times, but if he hits cross-court on one shot, on the next shot he can hit cross-court with .75 probability and down the line with .25 probability.

a. Write a transition matrix for this problem.	b. If the player hit the first shot cross-court, what is the probability that he will hit the third shot down the line?

5. The transition matrix for people voting for candidates from various political parties in an election year is given below. If a person votes for the candidate from one party in an election, that person may vote for the same party in the next election or may switch to vote for a candidate from another party in the next election. Democrats, Republicans, and Independents are denoted by the letters D, R, and I.

		Next Election D	R	I
This Election	D	.6	.3	.1
	R	.3	.6	.1
	I	.2	.2	.6

Assume there is an election every year so that the transition period is 1 year.

a. Find the probability that a person who votes Democratic in the current election will vote Republican in the next election.	b. Find the probability that a person who votes Democratic in the current election will vote Republican in the election two years from now.
c. Find the probability that a person who votes Republican in the current election will vote Independent in the election two years from now.	d. Find the probability that a person who votes Democratic in the current election will vote independent in the election three years from now.

This page titled 7.2.2.1: Introduction to Markov Chains (Exercises) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rupinder Sekhon and Roberta Bloom via source content that was edited to the style and standards of the LibreTexts platform.

- 10.1.1: Introduction to Markov Chains (Exercises) by Rupinder Sekhon and Roberta Bloom is licensed CC BY 4.0. Original source: https://www.deanza.edu/faculty/bloomroberta/math11/afm3files.html.html.

7.2.3: Regular Markov Chains

Learning Objectives

In this section, you will learn to:

1. Identify Regular Markov Chains, which have an equilibrium or steady state in the long run
2. Find the long term equilibrium for a Regular Markov Chain.

At the end of Section 10.1, we examined the transition matrix T for Professor Symons walking and biking to work. As we calculated higher and higher powers of T, the matrix started to stabilize, and finally it reached its **steady-state** or **state of equilibrium**. When that happened, all the row vectors became the same, and we called one such row vector a **fixed probability vector** or an **equilibrium vector** E. Furthermore, we discovered that ET = E.

In this section, we wish to answer the following four questions.

1. Does every Markov chain reach a state of equilibrium? Is there a way to determine if a Markov chain reaches a state of equilibrium?
2. Does the product of an equilibrium vector and its transition matrix always equal the equilibrium vector? That is, does ET = E?
3. Can the equilibrium vector E be found without raising the matrix to higher powers?
4. Does the long term market share distribution for a Markov chain depend on the initial market share?

Exercise 7.2.3.1

Does every Markov chain reach the state of equilibrium?

Answer available at the end of this section.

Example 7.2.3.1

Determine whether the following Markov chains are regular.

a. $\mathrm{A} = \begin{bmatrix} 0 & 1 \\ .4 & .6 \end{bmatrix}$

b. $\mathrm{B} = \begin{bmatrix} 1 & 0 \\ 3 & 7 \end{bmatrix}$

Solution

a.) The transition matrix A does not have all positive entries. But it is a regular Markov chain because

$$A^2 = \begin{bmatrix} .40 & .60 \\ .24 & .76 \end{bmatrix}$$

has only positive entries.

b.) The matrix B is not a regular Markov chain because every power of B has an entry 0 in the first row, second column position.

$$\mathrm{B} = \begin{bmatrix} 1 & 0 \\ .3 & .7 \end{bmatrix} \quad \text{and} \quad \mathrm{B}^2 = \begin{bmatrix} 1 & 0 \\ .3 & .7 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ .3 & .7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ .51 & .49 \end{bmatrix}$$

Since B is a 2×2 matrix, $m = (2-1)^2 + 1 = 2$. We've examined B and B^2, and discovered that neither has all positive entries. We don't need to examine any higher powers of B; B is not a regular Markov chain.

In fact, we can show that all 2 by 2 matrices that have a zero in the first row, second column position are not regular. Consider the following matrix M.

$$M = \begin{bmatrix} a & 0 \\ b & c \end{bmatrix}$$

$$M^2 = \begin{bmatrix} a & 0 \\ b & c \end{bmatrix} \begin{bmatrix} a & 0 \\ b & c \end{bmatrix} = \begin{bmatrix} a \cdot a + 0 \cdot b & a \cdot 0 + 0 \cdot c \\ b \cdot a + c \cdot b & b \cdot 0 + c \cdot c \end{bmatrix}$$

Observe that the first row, second column entry, $a \cdot 0 + 0 \cdot c$, will always be zero, regardless of what power we raise the matrix to.

Exercise 7.2.3.2

Does the product of an equilibrium vector and its transition matrix always equal the equilibrium vector? That is, does ET = E?

Answer available at the end of this section.

Exercise 7.2.3.3

Can the equilibrium vector E be found without raising the transition matrix T to large powers?

Answer available at the end of this section.

As a result of our work in Exercise 7.2.3.2 and 7.2.3.3, we see that we have a choice of methods to find the equilibrium vector.

Method 1: We can determine if the transition matrix T is regular. If T is regular, we know there is an equilibrium and we can use technology to find a high power of T.

- For the question of what is a sufficiently high power of T, there is no "exact" answer.
- Select a "high power", such as $n = 30$, or $n = 50$, or $n = 98$. Evaluate T^n on your calculator or with a computer. Check if $T^{n+1} = T^n$. If $T^{n+1} = T^n$ and all the rows of T^n are the same, then we've found the equilibrium. The equilibrium vector is one row of T^n. But if you did not find equilibrium yet for a regular Markov chain, try using a higher power of T.

Method 2: We can solve the matrix equation ET=E.

- The disadvantage of this method is that it is a bit harder, especially if the transition matrix is larger than 2×2. However it's not as hard as it seems, if T is not too large a matrix, because we can use the methods we learned in chapter 2 to solve the system of linear equations, rather than doing the algebra by hand.
- The advantage of solving ET = E as in Method 2 is that it can be used with matrices that are not regular. If a matrix is regular, it is guaranteed to have an equilibrium solution. If a matrix is not regular, then it may or may not have an equilibrium solution, and solving ET = E will allow us to prove that it has an equilibrium solution even if the matrix is not regular. (In mathematics we say that being a regular matrix is a "sufficient" condition for having an equilibrium, but is not a necessary condition.)

Exercise 7.2.3.4

Does the long term market share for a Markov chain depend on the initial market share?

Answer available at the end of this section.

Example 7.2.3.2

Three companies, A, B, and C, compete against each other. The transition matrix T for people switching each month among them is given by the following transition matrix.

Next Month

		Next Month		
		Company A	Company B	Company C
Initial Month	Company A	.1	.3	.6
	Company B	.6	.2	.2
	Company C	.1	.3	.6

If the initial market share for the companies A, B, and C is $\begin{bmatrix} .25 & .35 & .40 \end{bmatrix}$, what is the long term distribution?

Solution

Since the long term market share does not depend on the initial market share, we can simply raise the transition market share to a large power and get the distribution.

$$T^{20} = \begin{bmatrix} 13/55 & 3/11 & 27/55 \end{bmatrix}$$

In the long term, Company A has 13/55 (about 23.64%) of the market share, Company B has 3/11 (about 27.27%) of the market share, and Company C has 27/55 (about 49.09%) of the market share.

We summarize as follows:

Regular Markov Chains

A Markov chain is said to be a Regular Markov chain if some power of it has only positive entries.

Let T be a transition matrix for a regular Markov chain.

1. As we take higher powers of T, T^n, as n becomes large, approaches a state of equilibrium.
2. If V_0 is any distribution vector, and E an equilibrium vector, then $V_0T^n = E$.
3. Each row of the equilibrium matrix T^n is a unique equilibrium vector E such that ET = E.
4. The equilibrium distribution vector E can be found by letting ET = E.

Answers

Exercise 7.2.3.1

No. Some Markov chains reach a state of equilibrium but some do not. Some Markov chains transitions do not settle down to a fixed or equilibrium pattern. Therefore we'd like to have a way to identify Markov chains that do reach a state of equilibrium.

One type of Markov chains that do reach a state of equilibrium are called **regular** Markov chains. A Markov chain is said to be a **regular Markov chain** if some power of its transition matrix T has only positive entries.

To determine if a Markov chain is regular, we examine its transition matrix T and powers, T^n, of the transition matrix. If we find any power n for which T^n has only positive entries (no zero entries), then we know the Markov chain is regular and is guaranteed to reach a state of equilibrium in the long run.

Fortunately, we don't have to examine too many powers of the transition matrix T to determine if a Markov chain is regular; we use technology, calculators or computers, to do the calculations. There is a theorem that says that if an $n \times n$ transition matrix represents n states, then we need only examine powers T^m up to $m = (n-1)^2 + 1$.

If some power of the transition matrix T^m is going to have only positive entries, then that will occur for some power $m \leq (n-1)^2 + 1$.

For example, if T is a 3×3 transition matrix, then

$$m = (n-1)^2 + 1 = (3-1)^2 + 1 = 5.$$

- If we examine T, T^2, T^3, T^4 and T^5, and find that any of those matrices has only positive entries, then we know T is regular.
- If however, T, T^2, T^3, T^4 and T^5 all have at least one zero entry and none of them have all positive entries, then we can stop checking. All higher powers of T will also have at least one zero entry, and T will not be regular.

Exercise 7.2.3.2

At this point, the reader may have already guessed that the answer is yes if the transition matrix is a regular Markov chain. We try to illustrate with the following example from Section 10.1.

A city is served by two cable TV companies, BestTV and CableCast. Due to their aggressive sales tactics, each year 40% of BestTV customers switch to CableCast; the other 60% of BestTV customers stay with BestTV. On the other hand, 30% of the CableCast customers switch to Best RV and 70% of CableCast customers stay with CableCast.

The transition matrix is given below.

		Next Year	
		BestTV	CableCast
Initial Year	BestTV	.60	.40
	CableCast	.30	.70

If the initial market share for BestTV is 20% and for CableCast is 80%, we'd like to know the long term market share for each company.

Let matrix T denote the transition matrix for this Markov chain, and V_0 denote the matrix that represents the initial market share. Then V_0 and T are as follows:

$$V_0 = \begin{bmatrix} .20 & .80 \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} .60 & .40 \\ .30 & .70 \end{bmatrix}$$

Since each year people switch according to the transition matrix T, after one year the distribution for each company is as follows:

$$V_1 = V_0 T = \begin{bmatrix} .20 & .80 \end{bmatrix} \begin{bmatrix} .60 & .40 \\ .30 & .70 \end{bmatrix} = \begin{bmatrix} .36 & .64 \end{bmatrix}$$

After two years, the market share for each company is

$$V_2 = V_1 T = \begin{bmatrix} .36 & .64 \end{bmatrix} \begin{bmatrix} .60 & .40 \\ .30 & .70 \end{bmatrix} = \begin{bmatrix} .408 & .592 \end{bmatrix}$$

After three years the distribution is

$$V_3 = V_2 T = \begin{bmatrix} .408 & .592 \end{bmatrix} \begin{bmatrix} .60 & .40 \\ .30 & .70 \end{bmatrix} = \begin{bmatrix} .4224 & .5776 \end{bmatrix}$$

After 20 years the market share are given by $V_{20} = V_0 T^{20} = \begin{bmatrix} 3/7 & 4/7 \end{bmatrix}$.

After 21 years, $V_{21} = V_0 T^{21} = [3/7 \quad 4/7]$; market shares are stable and did not change.

The market share after 20 years has stabilized to $\begin{bmatrix} 3/7 & 4/7 \end{bmatrix}$. This means that

$$\begin{bmatrix} 3/7 & 4/7 \end{bmatrix} \begin{bmatrix} .60 & .40 \\ .30 & .70 \end{bmatrix} = \begin{bmatrix} 3/7 & 4/7 \end{bmatrix}$$

Once the market share reaches an equilibrium state, it stays the same, that is, ET = E.

This helps us answer the next question.

Exercise 7.2.3.3

The answer to the second question provides us with a way to find the equilibrium vector E. The answer lies in the fact that ET = E.

Since we have the matrix T, we can determine E from the statement ET = E.

Suppose $E = \begin{bmatrix} e & 1-e \end{bmatrix}$, then ET = E gives us

$$\begin{bmatrix} e & 1-e \end{bmatrix} \begin{bmatrix} .60 & .40 \\ .30 & .70 \end{bmatrix} = \begin{bmatrix} e & 1-e \end{bmatrix}$$

$$\begin{bmatrix} (.60)e + .30(1-e) & (.40)e + .70(1-e) \end{bmatrix} = \begin{bmatrix} e & 1-e \end{bmatrix}$$

$$\begin{bmatrix} .30e + .30 & -.30e + .70 \end{bmatrix} = \begin{bmatrix} e & 1-e \end{bmatrix}$$

$$.30e + .30 = e$$

$$e = 3/7$$

Therefore, $E = \begin{bmatrix} 3/7 & 4/7 \end{bmatrix}$

Exercise 7.2.3.4

We will show that the final market share distribution for a Markov chain does not depend upon the initial market share. In fact, one does not even need to know the initial market share distribution to find the long term distribution.

Furthermore, the final market share distribution can be found by simply raising the transition matrix to higher powers.

Consider the initial market share $V_0 = \begin{bmatrix} .20 & .80 \end{bmatrix}$, and the transition matrix $T = \begin{bmatrix} .60 & .40 \\ .30 & .70 \end{bmatrix}$ for BestTV and CableCast in the above example.

Recall we found T^n, for very large n, to be $\begin{bmatrix} 3/7 & 4/7 \\ 3/7 & 4/7 \end{bmatrix}$.

Using our calculators, we can easily verify that for sufficiently large n (we used $n = 30$),

$$V_0T^n = \begin{bmatrix} .20 & .80 \end{bmatrix} \begin{bmatrix} 3/7 & 4/7 \\ 3/7 & 4/7 \end{bmatrix} = \begin{bmatrix} 3/7 & 4/7 \end{bmatrix}$$

No matter what the initial market share, the product is $\begin{bmatrix} 3/7 & 4/7 \end{bmatrix}$. If instead the initial share is $W_0 = \begin{bmatrix} .10 & .90 \end{bmatrix}$, then for sufficiently large n

$$W_0T^n = \begin{bmatrix} .10 & .90 \end{bmatrix} \begin{bmatrix} 3/7 & 4/7 \\ 3/7 & 4/7 \end{bmatrix} = \begin{bmatrix} 3/7 & 4/7 \end{bmatrix}$$

For any distribution $A = \begin{bmatrix} a & 1-a \end{bmatrix}$, for example,

$$\begin{bmatrix} a & 1-a \end{bmatrix} \begin{bmatrix} 3/7 & 4/7 \\ 3/7 & 4/7 \end{bmatrix} = \begin{bmatrix} 3/7(a) + 3/7(1-a) & 4/7(a) + 4/7(1-a) \end{bmatrix}$$

$$= \begin{bmatrix} 3/7 & 4/7 \end{bmatrix}$$

It makes sense; the entry $3/7(a) + 3/7(1-a)$, for example, will always equal 3/7.

This page titled 7.2.3: Regular Markov Chains is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rupinder Sekhon and Roberta Bloom via source content that was edited to the style and standards of the LibreTexts platform.

- 10.3: Regular Markov Chains by Rupinder Sekhon and Roberta Bloom is licensed CC BY 4.0. Original source: https://www.deanza.edu/faculty/bloomroberta/math11/afm3files.html.html.

7.2.3.1: Regular Markov Chains (Exercises)

SECTION 10.3 PROBLEM SET: REGULAR MARKOV CHAINS

1. Determine whether the following matrices are regular Markov chains.

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

2. Company I and Company II compete against each other, and the transition matrix for people switching from Company I to Company II is given below.

		TO	
		Company I	Company II
FROM	Company I	.3	.7
	Company II	.8	.2

a. If the initial market share is 40% for Company I and 60% for Company II, what will the market share be after 3 transitions?	b. If this trend continues, what is the long range expectation for the market?

3. Suppose the transition matrix for a tennis player is as follows, where C denotes the cross-court shots and D denotes down-the-line shots.

		Next Shot	
		C	D
Previous Shot	C	.9	.1
	D	.7	.3

a. If the player hit the first shot cross-court, what is the probability he will hit the fourth shot cross-court?	b. Determine the long term shot distribution.

4. Professor Hay never orders eggs two days in a row, but if he orders tofu one day, then there is an equal probability that he will order tofu or eggs the next day.

Find the following:

a. If Professor Hay had eggs on Monday, what is the probability that he will have tofu on Friday?	b. Find the long term distribution for breakfast choices for Professor Hay.

5. In a bikeshare program with 3 bike stations, A, B, and C, people can borrow a bicycle at one station and return it to the same station or either of the other two stations. The transition matrix is:

$$T = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \end{array}$$

Find the following:

a. If a bicycle is initially at station A, what is the probability it will be at station C after 5 days?	b. If the initial distribution of bicycles is 50% at station A, 20% at station B, and 30% at station C, what will be the distribution after 2 days? After 5 days?
c. What will be the eventual long term distribution of bicycles at the stations?	d. If initially the distribution of bicycles at the stations was evenly distributed with one third of the bicycles at each station, will the eventual long term distribution be different than if the initial distribution is as given in part (c)?

6. Suppose that a country has 3 political parties: the Conservative (C), Liberal (L), and National (N) parties. If a person votes for the candidate from one party in an election, that person may decide to vote for the same party in the next election or may switch to vote for a candidate from another party in the next election. The transition matrix is:

$$\begin{array}{cc} & \text{NEXT ELECTION} \\ & \begin{array}{ccc} C & L & N \end{array} \\ \begin{array}{c} \text{THIS} \\ \text{ELECTION} \end{array} & \begin{array}{c} \\ C \\ L \\ N \end{array} \begin{bmatrix} .5 & .4 & .1 \\ .3 & .4 & .3 \\ .2 & .2 & .6 \end{bmatrix} \end{array}$$

Assume there is an election every year, so the transition time is one year. Find the following.

a. If a person voted for the Liberal party in this election, find the probability that the person votes for the National party in the next election.	b. If a person voted for the National party in this election, find the probability that the person votes for the Conservative party in the election two years from now.
c. If in this election Conservatives received 25% of the votes, Liberals 30% of the votes, and Nationals the remaining 45% of the votes, what is the predicted distribution for the next election?	d. Assuming the current distribution from part (c), what will the distribution be in the election two years from now?

e. Assuming the current distribution from part (c), what will the distribution be in the election three years from now?	f. Determine the long term distribution.

Question 7 refers to the following:

Markov chains play an important role in online search.

"PageRank is an algorithm used by Google Search to rank websites in their search engine results. PageRank was named after Larry Page, one of the founders of Google. PageRank is a way of measuring the importance of website pages"
Source: https://en.Wikipedia.org/wiki/PageRank under the Creative Commons Attribution-ShareAlike License;

The theory behind PageRank is that pages that are linked to more often are more important and useful; identifying those that are linked to more often about a topic helps identify the pages that should be presented as most pertinent in a search.

In real world search, there are thousands or millions of pages linking together, resulting in huge transition matrices. Because of the size and other properties of these matrices, the mathematics behind PageRank is more sophisticated than the small example we examine here with only four websites. However our example is adequate to convey the main concept of PageRank and its use in search algorithms.

It should be noted that real world search algorithms, PageRank or similar Markov chain ranking schemes are only one of a variety of methodsused.

Suppose we have 4 webpages that contains links to each other. We call the pages A, B, C, D.

- From page A, 30% of people link to page B, 50% link to page C, and 20% link to page D
- From page B, 50% of popele link to page A and 50% link to page D
- From page C, 10% of people link to page B, 70% link to page C, and 20% link to page D
- From page D, 20% of people link to page A, 40% to page B, 10% to page C, and 30% link to page D

(In this example, when a page links to itself, it means that a person viewing the page stays at that page and does not link to another page.)

a. Write the transition matrix.	b. T is a 4×4 matrix, $n = 4$ states. Use the formula $m = (n-1)^2 + 1$ to find the highest power m that we need to check to determine if T is a regular Markov chain.
c. Is this a regular Markov chain? Explain how you determined that.	d. Find the equilibrium vector and write a sentence summarizing the long term distribution of visits to these sites based on this model.
e. In the equilibrium vector, the state with the highest probability has the highest "Page-rank" and as the probabilities decrease, the ranking decreases. Indicate the order of the ranking, from highest Page rank to lowest Page rank, of these 4 pages.	

This page titled 7.2.3.1: Regular Markov Chains (Exercises) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Rupinder Sekhon and Roberta Bloom via source content that was edited to the style and standards of the LibreTexts platform.

- 10.3.1: Regular Markov Chains (Exercises) by Rupinder Sekhon and Roberta Bloom is licensed CC BY 4.0. Original source: https://www.deanza.edu/faculty/bloomroberta/math11/afm3files.html.html.

7.3: 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 7.3.1: 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.

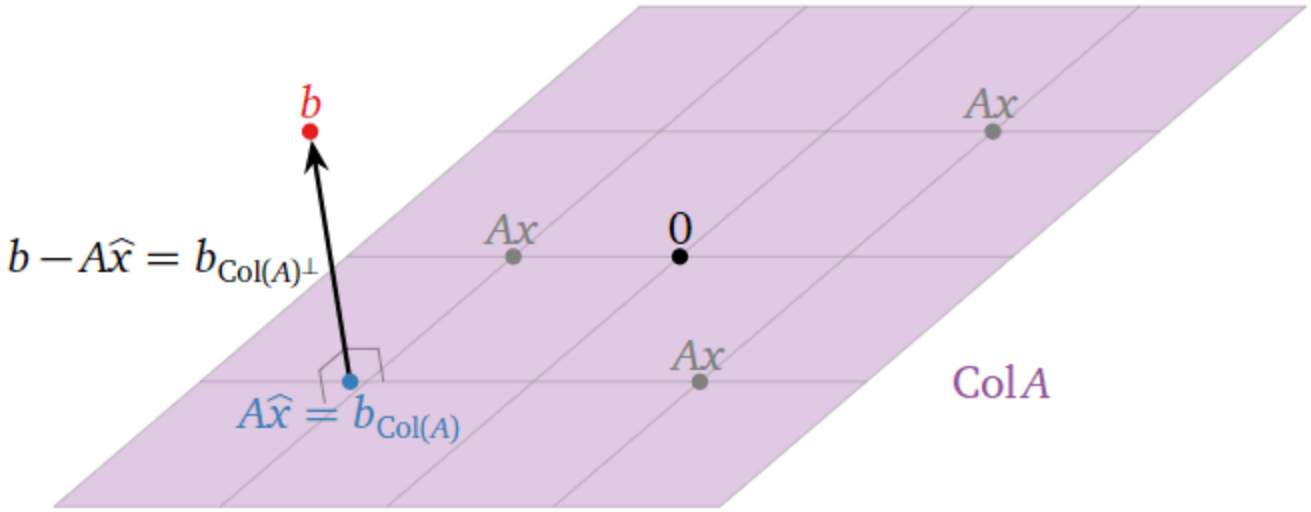


Figure 7.3.1

Note 7.3.1

A least-squares solution of $Ax = b$ is a solution $\hat{x}$ of the consistent equation $Ax = b_{\mathrm{Col}(A)}$

Note 7.3.2

If $Ax = b$ is consistent, then $b_{\mathrm{Col}(A)} = b$, 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 $b_{\mathrm{Col}(A)}$ with respect to the spanning set $\{v_1, v_2, \ldots, v_m\}$ of $\mathrm{Col}(A)$. (They are honest $\mathcal{B}$-coordinates if the columns of A are linearly independent.)

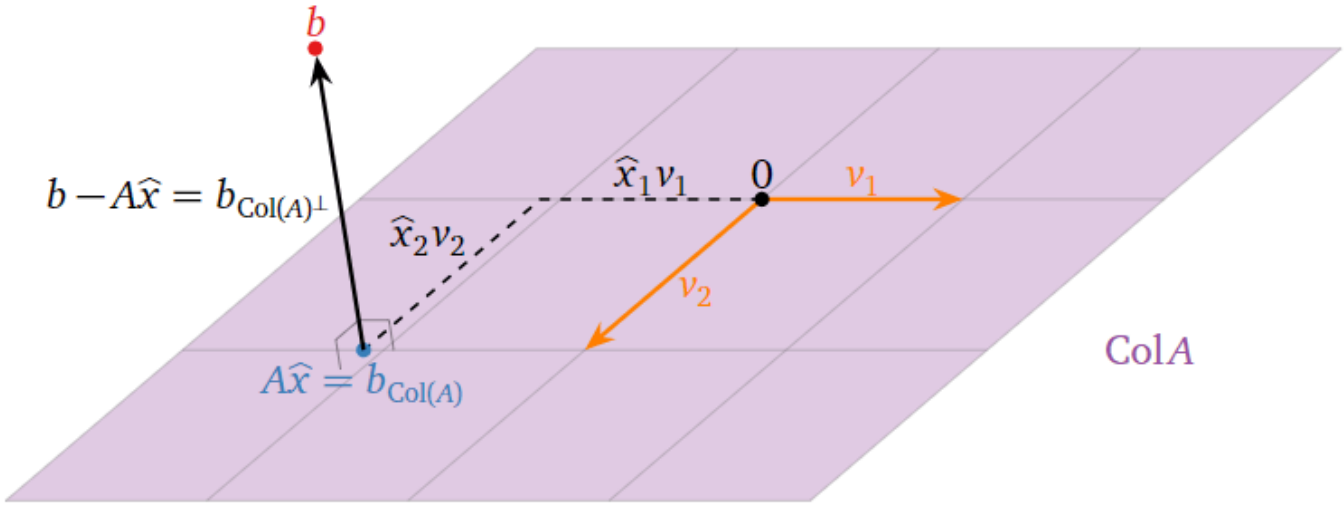


Figure 7.3.2

$$-1.10\,v_1 + 1.00\,v_2 = b_{\mathrm{Col}(A)}$$

$$A\hat{x} = \begin{bmatrix} 0.00 & 1.10 \\ 1.00 & 0.00 \\ 0.00 & -0.20 \end{bmatrix} \begin{bmatrix} -1.10 \\ 1.00 \end{bmatrix} = \begin{bmatrix} 1.10 \\ -1.10 \\ -0.20 \end{bmatrix} = b_{\mathrm{Col}(A)}$$

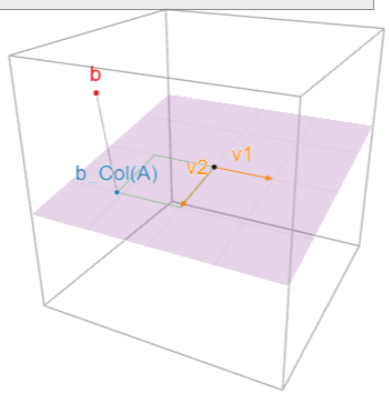


Figure 7.3.3: The violet plane is $\mathrm{Col}(A)$. The closest that Ax can get to b is the closest vector on $\mathrm{Col}(A)$ to b, which is the orthogonal projection $b_{\mathrm{Col}(A)}$ (in blue). The vectors v_1, v_2 are the columns of A, and the coefficients of $\hat{x}$ are the lengths of the green lines. Click and drag b to move it.

We learned to solve this kind of orthogonal projection problem in Section 6.3.

Theorem 7.3.1

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 available at the end of this section.

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^TA and the vector A^Tb.
2. Form the augmented matrix for the matrix equation $A^TAx = A^Tb$, 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 $b_{\text{Col}(A)}$ is equal to $A\hat{x}$.

Example 7.3.1

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

$$A^TA = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix}$$

and

$$A^Tb = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \end{pmatrix}.$$

We form an augmented matrix and row reduce:

$$\begin{pmatrix} 5 & 3 & 0 \\ 3 & 3 & 6 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & 5 \end{pmatrix}.$$

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 $b - A\hat{x} = b - b_{\text{Col}(A)} = b_{\text{Col}(A)^\perp}$. 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:

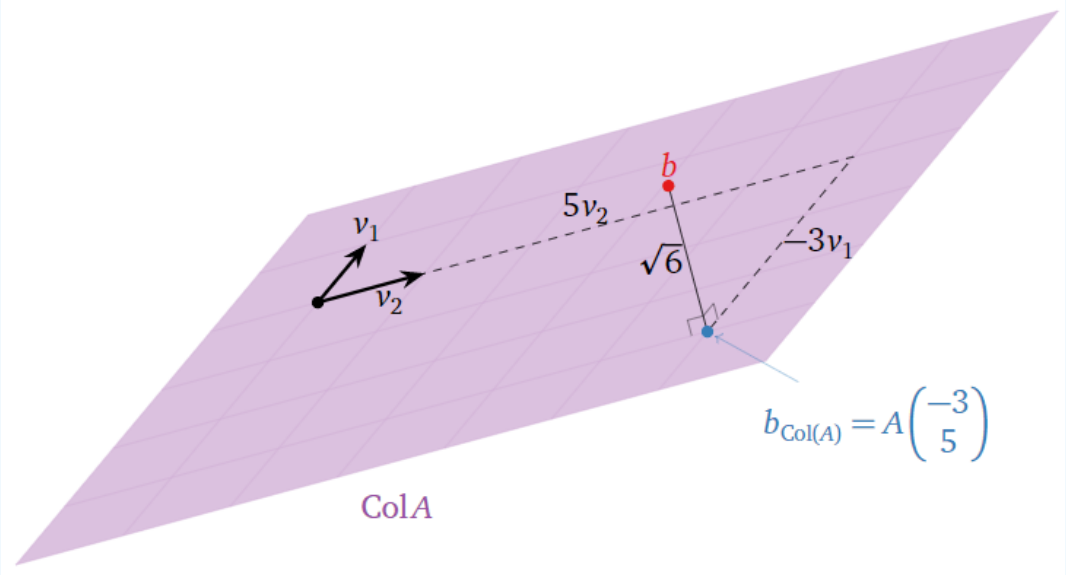


Figure 7.3.4

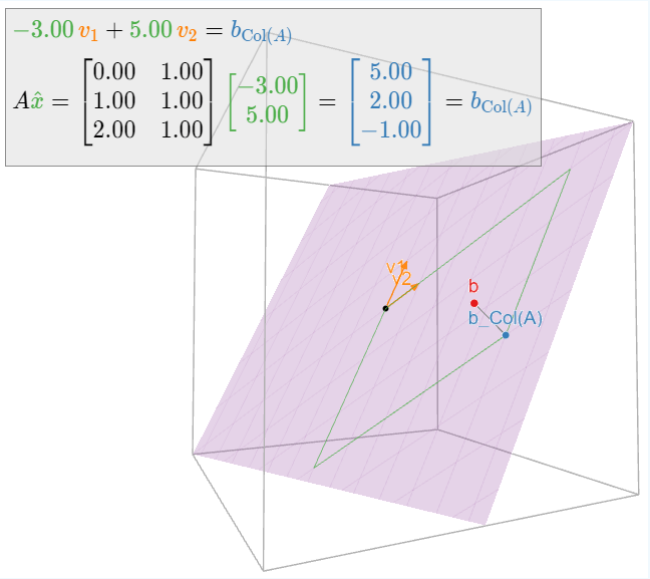


Figure 7.3.5: The violet plane is $\mathrm{Col}(A)$. The closest that Ax can get to b is the closest vector on $\mathrm{Col}(A)$ to b, which is the orthogonal projection $b_{\mathrm{Col}(A)}$ (in blue). The vectors v_1, v_2 are the columns of A, and the coefficients of $\hat{x}$ are the $\mathcal{B}$-coordinates of $b_{\mathrm{Col}(A)}$, where $\mathcal{B} = \{v_1, v_2\}$. lengths of the green lines.

Example 7.3.2

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:

$$\begin{pmatrix} 5 & -1 & 2 \\ -1 & 5 & -2 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 1/3 \\ 0 & 1 & -1/3 \end{pmatrix}.$$

Therefore, the only least-squares solution is $\hat{x} = \frac{1}{3}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

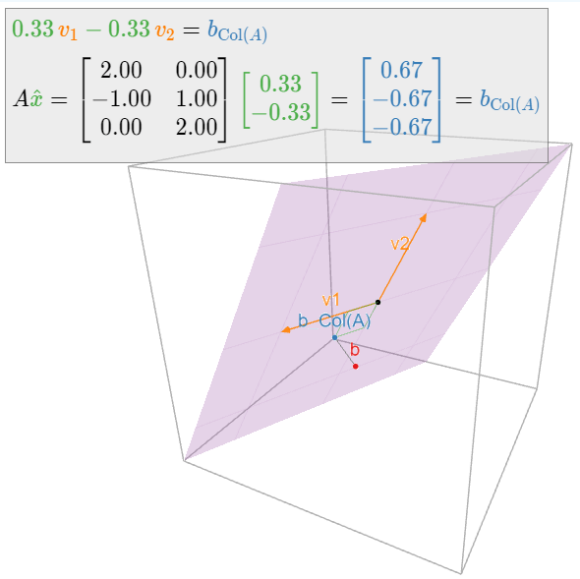


Figure 7.3.6: The red plane is $\mathrm{Col}(A)$. The closest that Ax can get to b is the closest vector on $\mathrm{Col}(A)$ to b, which is the orthogonal projection $b_{\mathrm{Col}(A)}$ (in blue). The vectors v_1, v_2 are the columns of A, and the coefficients of $\hat{x}$ are the $\mathcal{B}$-coordinates 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 7.3.2

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 available at the end of this section.

Example 7.3.3: 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:

$$\begin{pmatrix} 3 & 3 & -3 & 6 \\ 3 & 5 & -7 & 0 \\ -3 & -7 & 11 & 6 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 1 & 5 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

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:

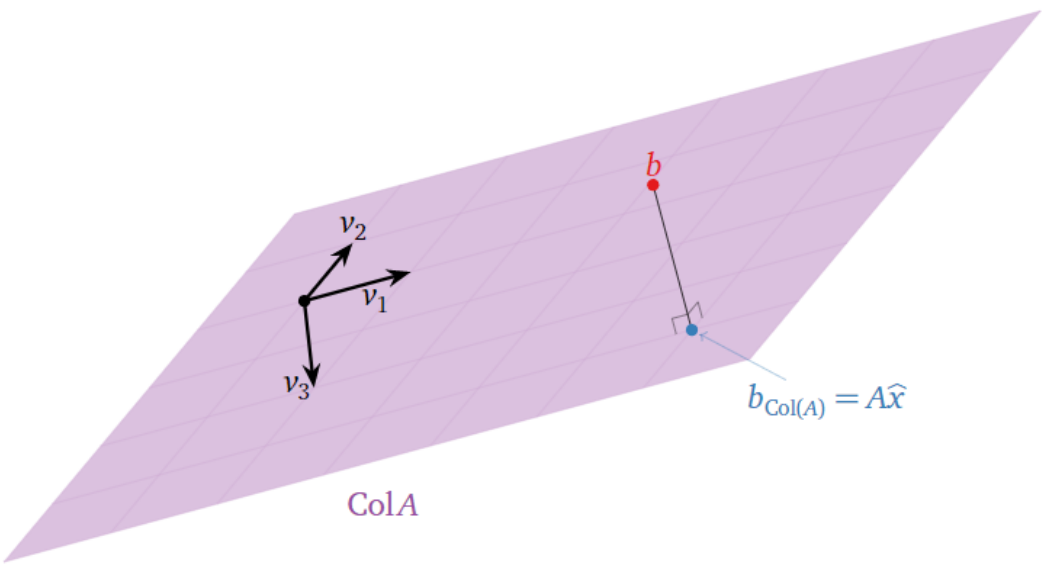


Figure 7.3.7

Therefore, there are many ways of writing $b_{\mathrm{Col}(A)}$ as a linear combination of v_1, v_2, v_3.

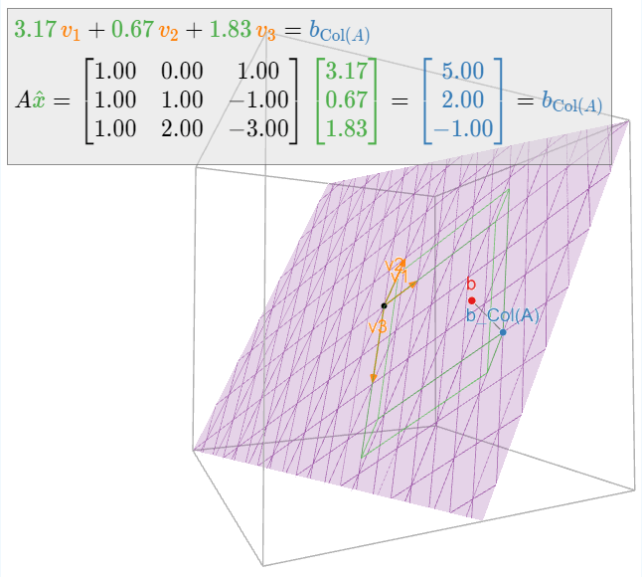


Figure 7.3.8: The three columns of A are coplanar, so there are many least-squares solutions. (The demo picks one solution when you move b.)

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 7.3.4

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 7.3.5: 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?

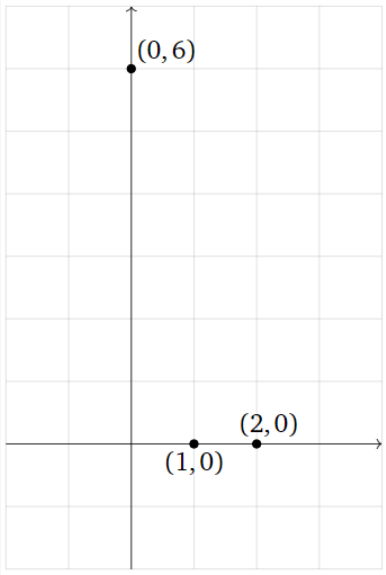


Figure 7.3.9

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:

$$\begin{aligned} 6 &= M \cdot 0 + B \\ 0 &= M \cdot 1 + B \\ 0 &= M \cdot 2 + B. \end{aligned} \tag{7.3.1}$$

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 7.3.1: the only least-squares solution to $Ax = b$ is $\hat{x} = \begin{pmatrix} M \\ B \end{pmatrix} = \begin{pmatrix} -3 \\ 5 \end{pmatrix}$, so the best-fit line is

$$y = -3x + 5.$$

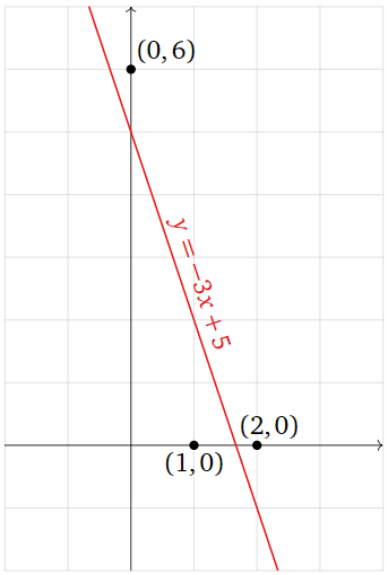


Figure 7.3.10

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 (7.3.1), 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:

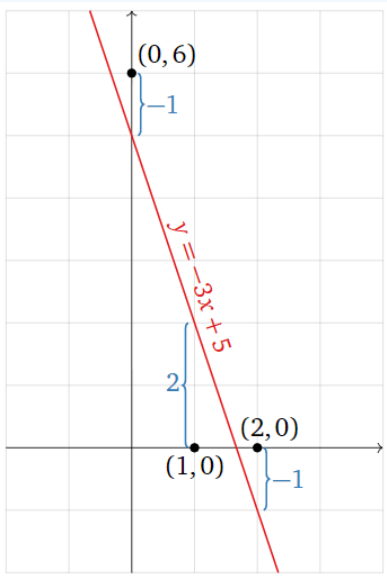


Figure 7.3.11

$$b - A\hat{x} = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} - A \begin{pmatrix} -3 \\ 5 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$$

The best-fit line minimizes the sum of the squares of these vertical distances.

Example 7.3.6: Interactive: Best-fit line

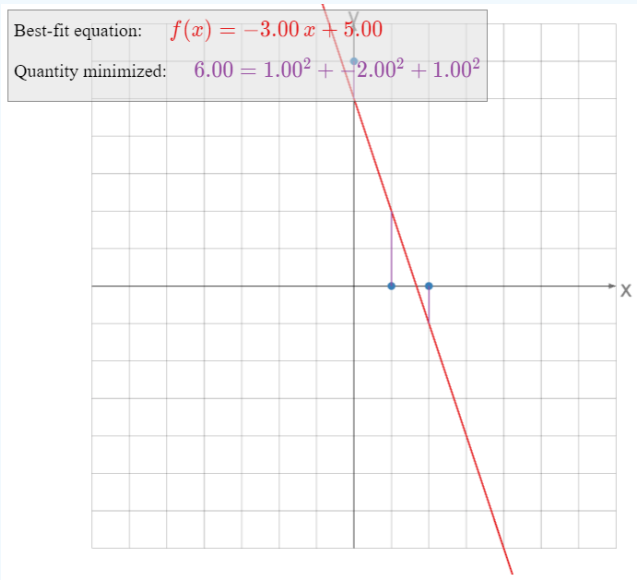


Figure 7.3.12: The best-fit line minimizes the sum of the squares of the vertical distances (violet). Click and drag the points to see how the best-fit line changes.

Example 7.3.7: Best-fit parabola

Find the parabola that best approximates the data points

$$(-1,\ 1/2), \quad (1,\ -1), \quad (2,\ -1/2), \quad (3,\ 2).$$

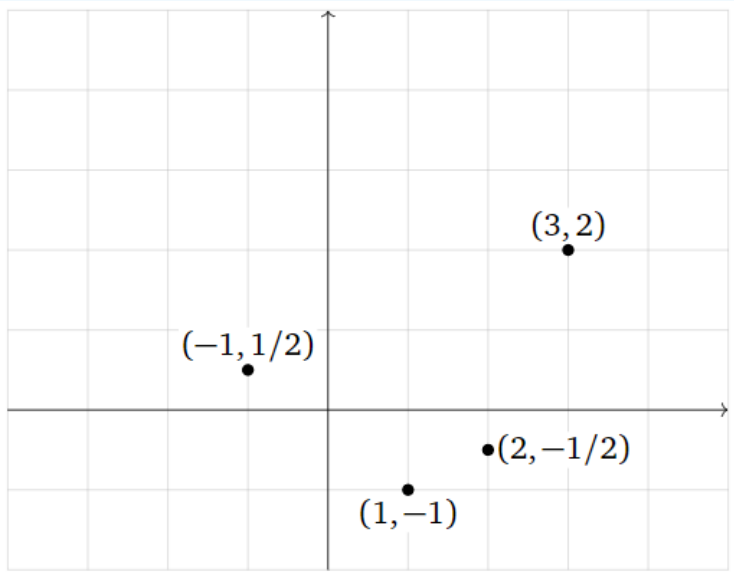


Figure 7.3.13

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:

$$\begin{aligned} \frac{1}{2} &= B(-1)^2 + C(-1) + D \\ -1 &= B(1)^2 + C(1) + D \\ -\frac{1}{2} &= B(2)^2 + C(2) + D \\ 2 &= B(3)^2 + C(3) + D. \end{aligned} \tag{7.3.2}$$

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:

$$\begin{pmatrix} 99 & 35 & 15 & 31/2 \\ 35 & 15 & 5 & 7/2 \\ 15 & 5 & 4 & 1 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 0 & 53/88 \\ 0 & 1 & 0 & -379/440 \\ 0 & 0 & 1 & -41/44 \end{pmatrix} \implies \hat{x} = \begin{pmatrix} 53/88 \\ -379/440 \\ -41/44 \end{pmatrix}.$$

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

$$88y = 53x^2 - \frac{379}{5}x - 82.$$

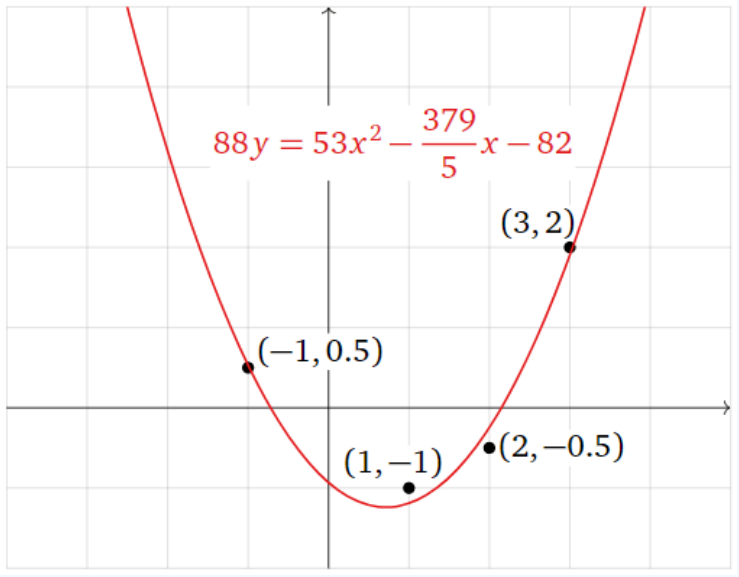


Figure 7.3.14

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 (7.3.2), 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:

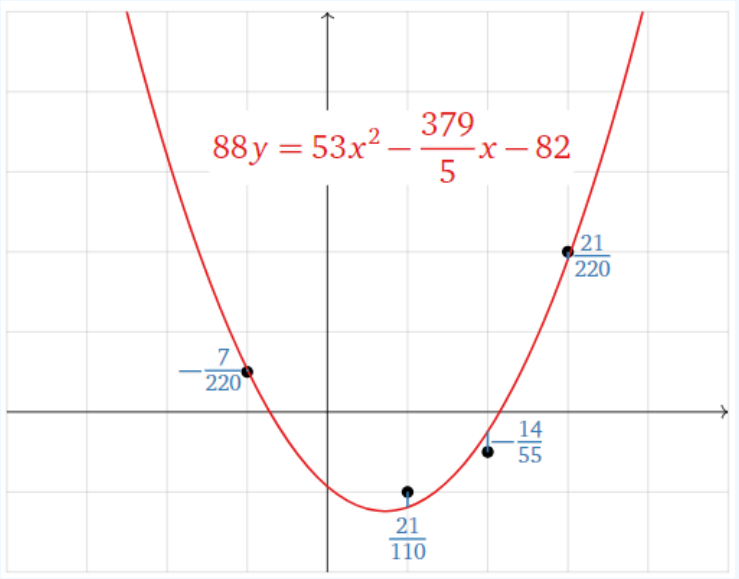


Figure 7.3.15

$$b - A\hat{x} = \begin{pmatrix} 1/2 \\ -1 \\ -1/2 \\ 2 \end{pmatrix} - A \begin{pmatrix} 53/88 \\ -379/440 \\ -41/44 \end{pmatrix} = \begin{pmatrix} -7/220 \\ 21/110 \\ -14/55 \\ 21/220 \end{pmatrix}$$

The best-fit parabola minimizes the sum of the squares of these vertical distances.

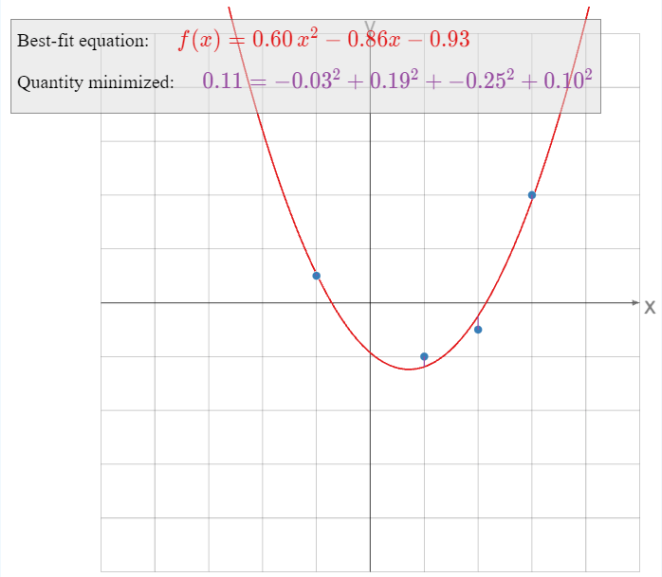


Figure 7.3.16: The best-fit parabola minimizes the sum of the squares of the vertical distances (violet). Click and drag the points to see how the best-fit parabola changes.

Example 7.3.8: 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:

$$\begin{aligned} B(1) + C(0) + D &= 0 \\ B(0) + C(1) + D &= 1 \\ B(-1) + C(0) + D &= 3 \\ B(0) + C(-1) + D &= 4 \end{aligned} \tag{7.3.3}$$

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 $A^T A x = A^T b$:

$$\left(\begin{array}{cccc} 2 & 0 & 0 & -3 \\ 0 & 2 & 0 & -3 \\ 0 & 0 & 4 & 8 \end{array} \right) \xrightarrow{\text{RREF}} \left(\begin{array}{cccc} 1 & 0 & 0 & -3/2 \\ 0 & 1 & 0 & -3/2 \\ 0 & 0 & 1 & 2 \end{array} \right) \implies \hat{x} = \begin{pmatrix} -3/2 \\ -3/2 \\ 2 \end{pmatrix}.$$

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 $f(x, y)$:

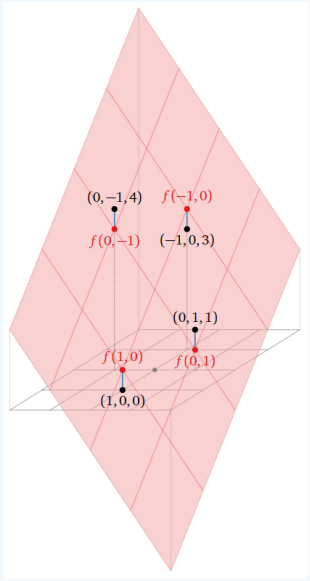


Figure 7.3.17

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 (7.3.3), 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.

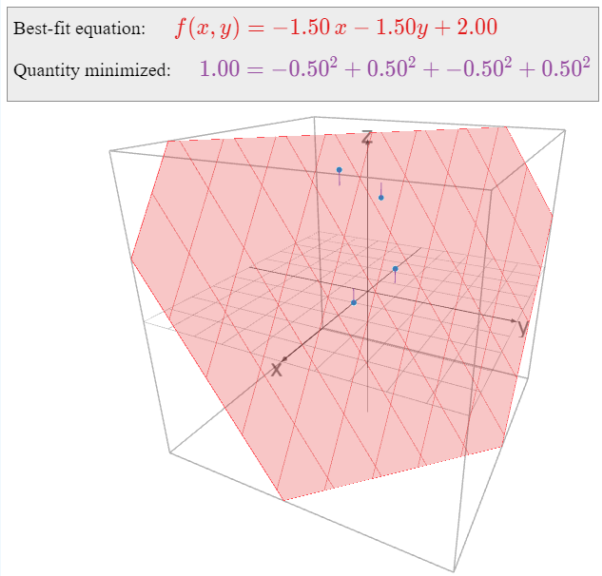


Figure 7.3.18: The best-fit linear function minimizes the sum of the squares of the vertical distances (violet). Click and drag the points to see how the best-fit linear function changes.

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 7.3.9: 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

$$\begin{pmatrix} -4 \\ -1 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ -1.5 \end{pmatrix}, \begin{pmatrix} -1 \\ .5 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ -.5 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ -1 \end{pmatrix}?$$

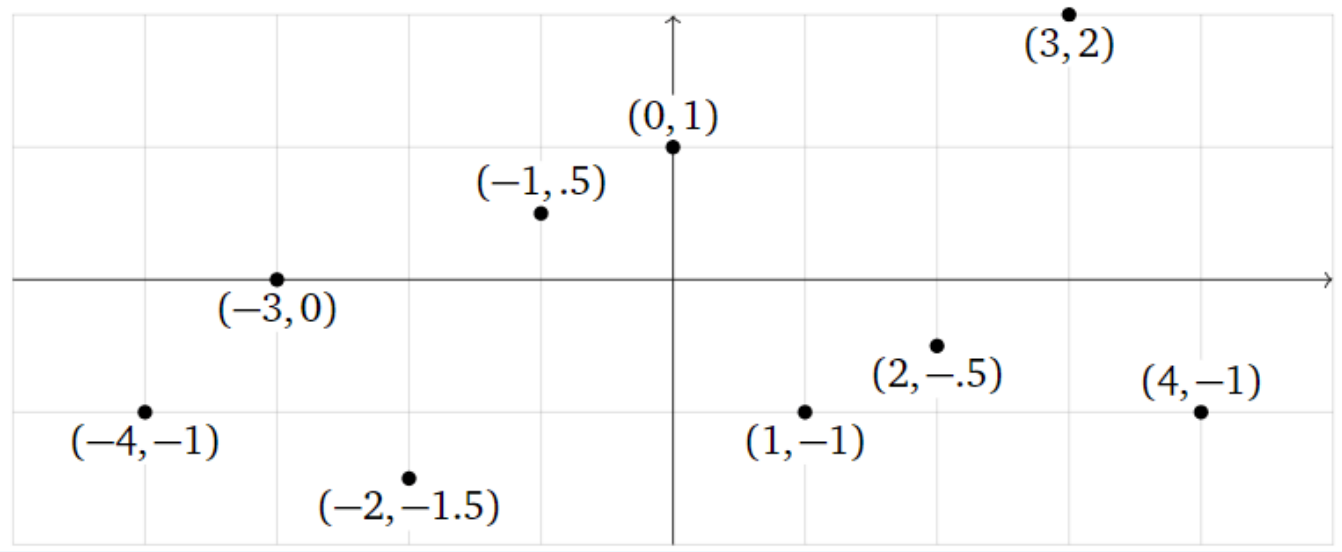


Figure 7.3.19

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

$$y \approx -0.1435 + 0.2611\cos(x) - 0.2337\sin(x) + 1.116\cos(2x) - 0.5997\sin(2x) \\ -0.2767\cos(3x) + 0.1076\sin(3x).$$

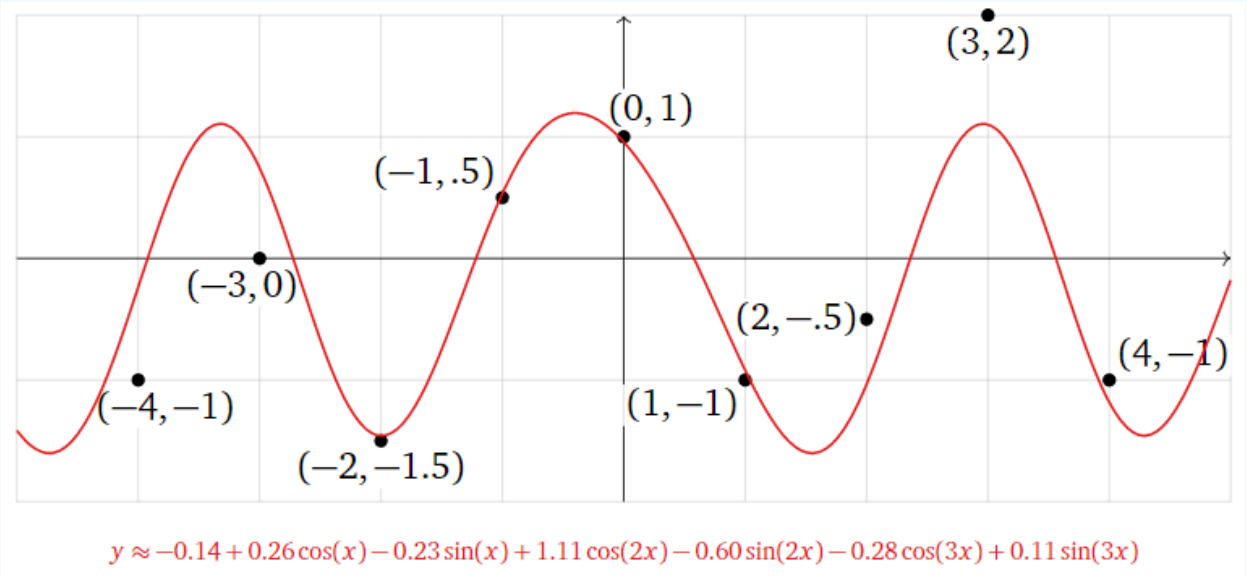


Figure 7.3.20

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.

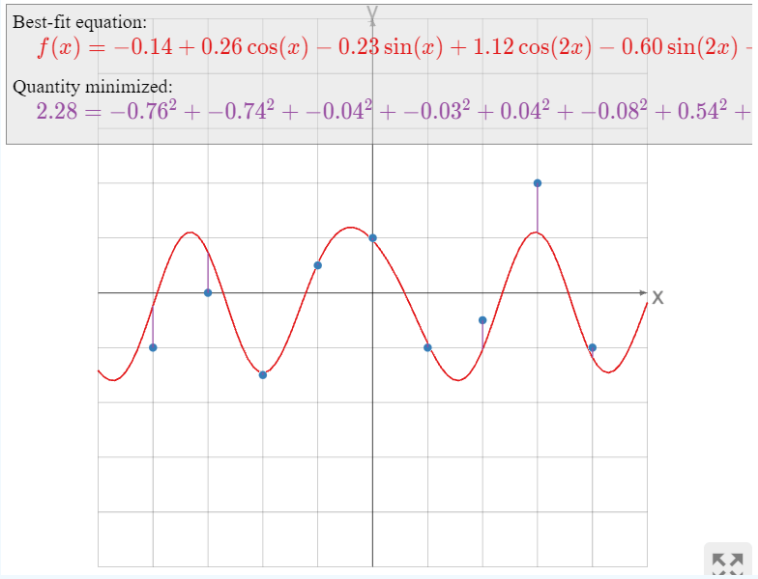


Figure 7.3.21: The best-fit function minimizes the sum of the squares of the vertical distances (violet). Click and drag the points to see how the best-fit function changes.

The next example has a somewhat different flavor from the previous ones.

Example 7.3.10: Best-fit ellipse

Find the best-fit ellipse through the points

$$(0,2),\ (2,1),\ (1,-1),\ (-1,-2),\ (-3,1),\ (-1,-1).$$

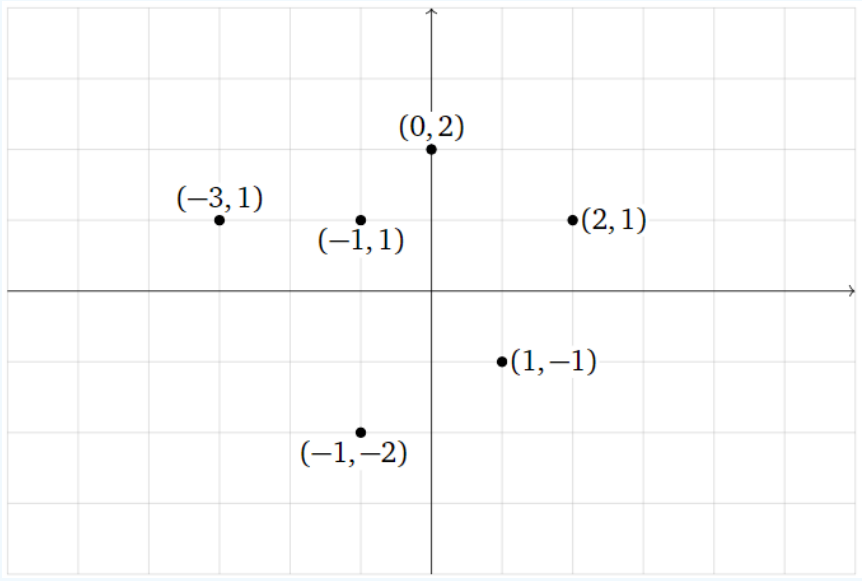


Figure 7.3.22

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:

$$\begin{array}{rcrcrcrcrcl} (0)^2 & + & B(2)^2 & + & C(0)(2) & + & D(0) & + & E(2) & + & F & = & 0 \\ (2)^2 & + & B(1)^2 & + & C(2)(1) & + & D(2) & + & E(1) & + & F & = & 0 \\ (1)^2 & + & B(-1)^2 & + & C(1)(-1) & + & D(1) & + & E(-1) & + & F & = & 0 \\ (-1)^2 & + & B(-2)^2 & + & C(-1)(2) & + & D(-1) & + & E(-2) & + & F & = & 0 \\ (-3)^2 & + & B(1)^2 & + & C(-3)(1) & + & D(-3) & + & E(1) & + & F & = & 0 \\ (-1)^2 & + & B(-1)^2 & + & C(-1)(-1) & + & D(-1) & + & E(-1) & + & F & = & 0. \end{array} \tag{7.3.4}$$

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

$$A^TA = \begin{pmatrix} 36 & 7 & -5 & 0 & 12 \\ 7 & 19 & 9 & -5 & 1 \\ -5 & 9 & 16 & 1 & -2 \\ 0 & -5 & 1 & 12 & 0 \\ 12 & 1 & -2 & 0 & 6 \end{pmatrix} \qquad A^Tb = \begin{pmatrix} -19 \\ 17 \\ 20 \\ -9 \\ -16 \end{pmatrix}.$$

We form an augmented matrix and row reduce:

$$\begin{pmatrix} 36 & 7 & -5 & 0 & 12 & -19 \\ 7 & 19 & 9 & -5 & 1 & 17 \\ -5 & 9 & 16 & 1 & -2 & 20 \\ 0 & -5 & 1 & 12 & 0 & -9 \\ 12 & 1 & -2 & 0 & 6 & -16 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 405/266 \\ 0 & 1 & 0 & 0 & 0 & -89/133 \\ 0 & 0 & 1 & 0 & 0 & 201/133 \\ 0 & 0 & 0 & 1 & 0 & -123/266 \\ 0 & 0 & 0 & 0 & 1 & -687/133 \end{pmatrix}.$$

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

$$266x^2 + 405y^2 - 178xy + 402x - 123y - 1374 = 0.$$

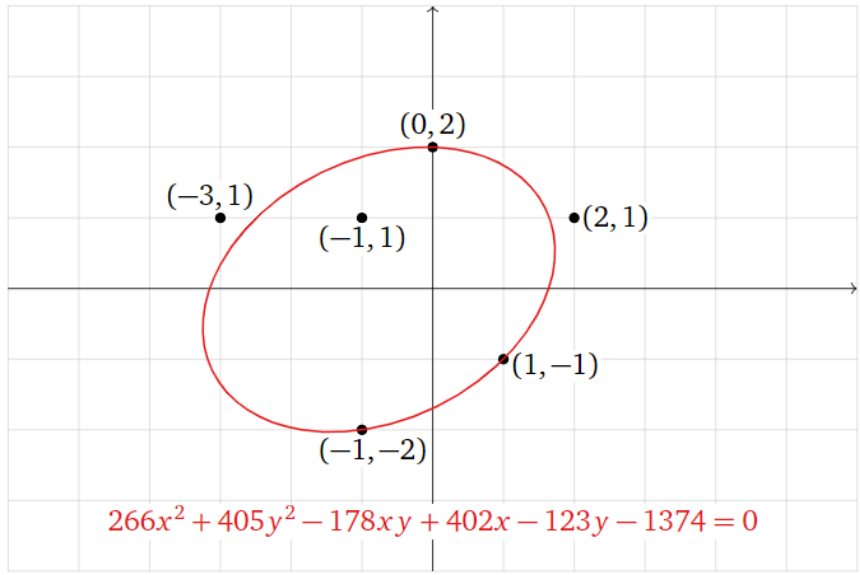


Figure 7.3.23

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 (7.3.4), 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 (7.3.4). 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 7.3.8 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 $(0,2,0)$, $(2,1,0)$, ... , which lie on the xy-plane, to the graph of $f(x,y)$.

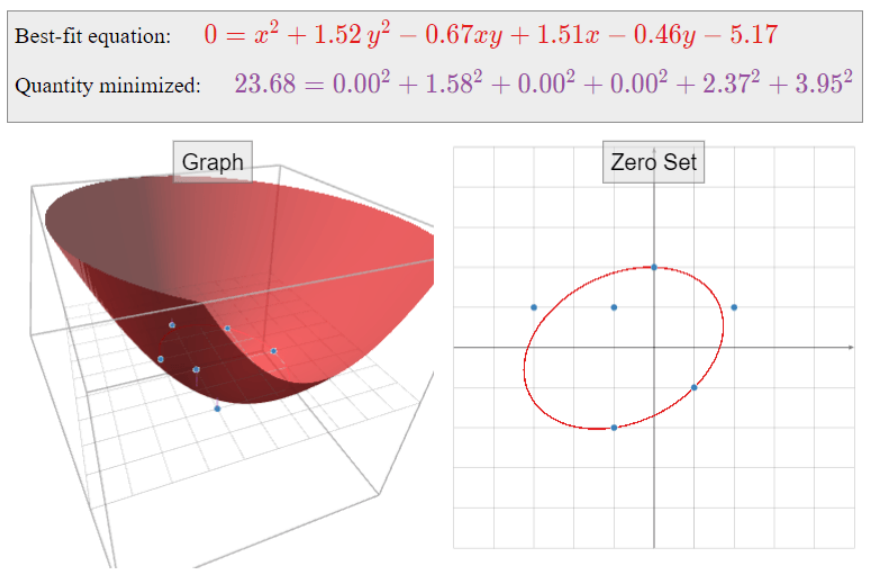


Figure 7.3.24: The best-fit ellipse minimizes the sum of the squares of the vertical distances (violet) from the points $(x,y,0)$ to the graph of $f(x,y)$ on the left. The ellipse itself is the zero set of $f(x,y)$, on the right. Click and drag the points on the right to see how the best-fit ellipse changes. Can you arrange the points so that the best-fit conic section is actually a hyperbola?

Note 7.3.3

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.

Answers

Theorem 7.3.1

By Theorem 6.3.2 in Section 6.3, if $\hat{x}$ is a solution of the matrix equation $A^TAx = A^Tb$, 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 $Ax = b_{\mathrm{Col}(A)}$.

Theorem 7.3.2

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 $Ax = b_{\mathrm{Col}(A)}$, 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.

This page titled 7.3: The Method of Least Squares is shared under a GNU FDL license and was authored, remixed, and/or curated by Dan Margalit and Joseph Rabinoff via source content that was edited to the style and standards of the LibreTexts platform.

- 6.5: The Method of Least Squares by Dan Margalit and Joseph Rabinoff is licensed GNU FDL. Original source: https://textbooks.math.gatech.edu/ila.

7.4: An Application to Linear Recurrences

It often happens that a problem can be solved by finding a sequence of numbers $x_0, x_1, x_2, \ldots$ where the first few are known, and subsequent numbers are given in terms of earlier ones. Here is a combinatorial example where the object is to count the number of ways to do something.

Example 7.4.1

An urban planner wants to determine the number x_k of ways that a row of k parking spaces can be filled with cars and trucks if trucks take up two spaces each. Find the first few values of x_k.

Solution

Clearly, $x_0 = 1$ and $x_1 = 1$, while $x_2 = 2$ since there can be two cars or one truck. We have $x_3 = 3$ (the 3 configurations are *ccc*, *cT*, and *Tc*) and $x_4 = 5$ (*cccc, ccT, cTc, Tcc,* and *TT*). The key to this method is to find a way to express each subsequent x_k in terms of earlier values. In this case we claim that

$$x_{k+2} = x_k + x_{k+1} \text{ for every } k \geq 0 \tag{7.4.1}$$

Indeed, every way to fill $k+2$ spaces falls into one of two categories: Either a car is parked in the first space (and the remaining $k+1$ spaces are filled in x_{k+1} ways), or a truck is parked in the first two spaces (with the other k spaces filled in x_k ways). Hence, there are $x_{k+1} + x_k$ ways to fill the $k+2$ spaces. This is Equation 7.4.1.

The recurrence in Equation 7.4.1 determines x_k for every $k \geq 2$ since x_0 and x_1 are given. In fact, the first few values are

$$\begin{array}{rcccc} x_0 & = & 1 & & \\ x_1 & = & 1 & & \\ x_2 & = & x_0 + x_1 & = & 2 \\ x_3 & = & x_1 + x_2 & = & 3 \\ x_4 & = & x_2 + x_3 & = & 5 \\ x_5 & = & x_3 + x_4 & = & 8 \\ \vdots & & \vdots & & \vdots \end{array}$$

Clearly, we can find x_k for any value of k, but one wishes for a "formula" for x_k as a function of k. It turns out that such a formula can be found using diagonalization. We will return to this example later.

A sequence $x_0, x_1, x_2, \ldots$ of numbers is said to be given **recursively** if each number in the sequence is completely determined by those that come before it. Such sequences arise frequently in mathematics and computer science, and also occur in other parts of science. The formula $x_{k+2} = x_{k+1} + x_k$ in Example 7.4.1 is an example of a **linear recurrence relation** of length 2 because x_{k+2} is the sum of the two preceding terms x_{k+1} and x_k; in general, the **length** is m if x_{k+m} is a sum of multiples of $x_k, x_{k+1}, \ldots, x_{k+m-1}$.

The simplest linear recursive sequences are of length 1, that is x_{k+1} is a fixed multiple of x_k for each k, say $x_{k+1} = ax_k$. If x_0 is specified, then $x_1 = ax_0, x_2 = ax_1 = a^2x_0$, and $x_3 = ax_2 = a^3x_0, \ldots$. Continuing, we obtain $x_k = a^kx_0$ for each $k \geq 0$, which is an explicit formula for x_k as a function of k (when x_0 is given).

Such formulas are not always so easy to find for all choices of the initial values. Here is an example where diagonalization helps.

Example 7.4.2

Suppose the numbers $x_0, x_1, x_2, \ldots$ are given by the linear recurrence relation

$$x_{k+2} = x_{k+1} + 6x_k \text{ for } k \geq 0$$

where x_0 and x_1 are specified. Find a formula for x_k when $x_0 = 1$ and $x_1 = 3$, and also when $x_0 = 1$ and $x_1 = 1$.

Solution

If $x_0 = 1$ and $x_1 = 3$, then

$$x_2 = x_1 + 6x_0 = 9, \quad x_3 = x_2 + 6x_1 = 27, \quad x_4 = x_3 + 6x_2 = 81$$

and it is apparent that

$$x_k = 3^k \text{ for } k = 0, 1, 2, 3, \text{ and } 4$$

This formula holds for all k because it is true for $k = 0$ and $k = 1$, and it satisfies the recurrence $x_{k+2} = x_{k+1} + 6x_k$ for each k as is readily checked.

However, if we begin instead with $x_0 = 1$ and $x_1 = 1$, the sequence continues

$$x_2 = 7, \quad x_3 = 13, \quad x_4 = 55, \quad x_5 = 133, \quad \ldots$$

In this case, the sequence is uniquely determined but no formula is apparent. Nonetheless, a simple device transforms the recurrence into a matrix recurrence to which our diagonalization techniques apply.

The idea is to compute the sequence $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \ldots$ of columns instead of the numbers $x_0, x_1, x_2, \ldots$, where

$$\mathbf{v}_k = \begin{bmatrix} x_k \\ x_{k+1} \end{bmatrix} \text{ for each } k \geq 0$$

Then $\mathbf{v}_0 = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is specified, and the numerical recurrence $x_{k+2} = x_{k+1} + 6x_k$ transforms into a matrix recurrence as follows:

$$\mathbf{v}_{k+1} = \begin{bmatrix} x_{k+1} \\ x_{k+2} \end{bmatrix} = \begin{bmatrix} x_{k+1} \\ 6x_k + x_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 6 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ x_{k+1} \end{bmatrix} = A\mathbf{v}_k$$

where $A = \begin{bmatrix} 0 & 1 \\ 6 & 1 \end{bmatrix}$. Thus these columns $\mathbf{v}_k$ are a linear dynamical system, so Theorem 3.5.1 applies provided the matrix A is diagonalizable.

We have $c_A(x) = (x - 3)(x + 2)$ so the eigenvalues are $\lambda_1 = 3$ and $\lambda_2 = -2$ with corresponding eigenvectors $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ as the reader can check. Since $P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 3 & 2 \end{bmatrix}$ is invertible, it is a diagonalizing matrix for A. The coefficients b_i in Theorem 3.5.1 are given by $\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = P^{-1}\mathbf{v}_0 = \begin{bmatrix} \frac{3}{5} \\ \frac{-2}{5} \end{bmatrix}$, so that the theorem gives

$$\begin{bmatrix} x_k \\ x_{k+1} \end{bmatrix} = \mathbf{v}_k = b_1\lambda_1^k\mathbf{x}_1 + b_2\lambda_2^k\mathbf{x}_2 = \frac{3}{5}3^k \begin{bmatrix} 1 \\ 3 \end{bmatrix} + \frac{-2}{5}(-2)^k \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

Equating top entries yields

$$x_k = \frac{1}{5}\left[3^{k+1} - (-2)^{k+1}\right] \text{ for } k \geq 0$$

This gives $x_0 = 1 = x_1$, and it satisfies the recurrence $x_{k+2} = x_{k+1} + 6x_k$ as is easily verified. Hence, it is the desired formula for the x_k.

Returning to Example 7.4.1, these methods give an exact formula and a good approximation for the numbers x_k in that problem.

Example 7.4.3

In Example 7.4.1, an urban planner wants to determine x_k, the number of ways that a row of k parking spaces can be filled with cars and trucks if trucks take up two spaces each. Find a formula for x_k and estimate it for large k.

Solution

We saw in Example 7.4.1 that the numbers x_k satisfy a linear recurrence

$$x_{k+2} = x_k + x_{k+1} \text{ for every } k \geq 0$$

If we write $\mathbf{v}_k = \begin{bmatrix} x_k \\ x_{k+1} \end{bmatrix}$ as before, this recurrence becomes a matrix recurrence for the $\mathbf{v}_k$:

$$\mathbf{v}_{k+1} = \begin{bmatrix} x_{k+1} \\ x_{k+2} \end{bmatrix} = \begin{bmatrix} x_{k+1} \\ x_k + x_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ x_{k+1} \end{bmatrix} = A\mathbf{v}_k$$

for all $k \geq 0$ where $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$. Moreover, A is diagonalizable here. The characteristic polynomial is $c_A(x) = x^2 - x - 1$ with roots $\frac{1}{2}\left[1 \pm \sqrt{5}\right]$ by the quadratic formula, so A has eigenvalues

$$\lambda_1 = \frac{1}{2}\left(1 + \sqrt{5}\right) \text{ and } \lambda_2 = \frac{1}{2}\left(1 - \sqrt{5}\right)$$

Corresponding eigenvectors are $\mathbf{x}_1 = \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix}$ respectively as the reader can verify. As the matrix $P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix}$ is invertible, it is a diagonalizing matrix for A. We compute the coefficients b_1 and b_2 (in Theorem 3.5.1) as follows:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = P^{-1}\mathbf{v}_0 = \frac{1}{-\sqrt{5}} \begin{bmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} \lambda_1 \\ -\lambda_2 \end{bmatrix}$$

where we used the fact that $\lambda_1 + \lambda_2 = 1$. Thus Theorem 3.5.1 gives

$$\begin{bmatrix} x_k \\ x_{k+1} \end{bmatrix} = \mathbf{v}_k = b_1\lambda_1^k\mathbf{x}_1 + b_2\lambda_2^k\mathbf{x}_2 = \frac{\lambda_1}{\sqrt{5}}\lambda_1^k \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix} - \frac{\lambda_2}{\sqrt{5}}\lambda_2^k \begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix}$$

Comparing top entries gives an exact formula for the numbers x_k:

$$x_k = \frac{1}{\sqrt{5}}\left[\lambda_1^{k+1} - \lambda_2^{k+1}\right] \text{ for } k \geq 0$$

Finally, observe that λ_1 is dominant here (in fact, $\lambda_1 = 1.618$ and $\lambda_2 = -0.618$ to three decimal places) so λ_2^{k+1} is negligible compared with λ_1^{k+1} is large. Thus,

$$x_k \approx \frac{1}{\sqrt{5}}\lambda_1^{k+1} \text{ for each } k \geq 0.$$

This is a good approximation, even for as small a value as $k = 12$. Indeed, repeated use of the recurrence $x_{k+2} = x_k + x_{k+1}$ gives the exact value $x_{12} = 233$, while the approximation is $x_{12} \approx \frac{(1.618)^{13}}{\sqrt{5}} = 232.94$.

The sequence $x_0, x_1, x_2, \ldots$ in Example 3.6.3 was first discussed in 1202 by Leonardo Pisano of Pisa, also known as Fibonacci,[1] and is now called the **Fibonacci sequence**. It is completely determined by the conditions $x_0 = 1$, $x_1 = 1$ and the recurrence $x_{k+2} = x_k + x_{k+1}$ for each $k \geq 0$. These numbers have been studied for centuries and have many interesting properties (there is even a journal, the *Fibonacci Quarterly*, devoted exclusively to them). For example, biologists have discovered that the arrangement of leaves around the stems of some plants follow a Fibonacci pattern. The formula $x_k = \frac{1}{\sqrt{5}}\left[\lambda_1^{k+1} - \lambda_2^{k+1}\right]$ in Example 3.6.3 is called the **Binet formula**. It is remarkable in that the x_k are integers but λ_1 and λ_2 are not. This phenomenon can occur even if the

eigenvalues λ_i are nonreal complex numbers.

We conclude with an example showing that *nonlinear* recurrences can be very complicated.

Example 7.4.4

Suppose a sequence $x_0, x_1, x_2, \ldots$ satisfies the following recurrence:

$$x_{k+1} = \begin{cases} \frac{1}{2}x_k & \text{if } x_k \text{ is even} \\ 3x_k + 1 & \text{if } x_k \text{ is odd} \end{cases}$$

If $x_0 = 1$, the sequence is $1, 4, 2, 1, 4, 2, 1, \ldots$ and so continues to cycle indefinitely. The same thing happens if $x_0 = 7$. Then the sequence is

$$7, 22, 11, 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1, \ldots$$

and it again cycles. However, it is not known whether every choice of x_0 will lead eventually to 1. It is quite possible that, for some x_0, the sequence will continue to produce different values indefinitely, or will repeat a value and cycle without reaching 1. No one knows for sure.

---

1. Fibonacci was born in Italy. As a young man he travelled to India where he encountered the "Fibonacci" sequence. He returned to Italy and published this in his book *Liber Abaci* in 1202. In the book he is the first to bring the Hindu decimal system for representing numbers to Europe.

This page titled 7.4: An Application to Linear Recurrences is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by W. Keith Nicholson via source content that was edited to the style and standards of the LibreTexts platform.

- 3.6: An Application to Linear Recurrences by W. Keith Nicholson is licensed CC BY-NC-SA 4.0. Original source: https://lyryx.com/linear-algebra-applications.

7.4E: An Application to Linear Recurrences Exercises

Exercise $7.4E.1$

Solve the following linear recurrences.

a. $x_{k+2} = 3x_k + 2x_{k+1}$, where $x_0 = 1$ and $x_1 = 1$.
b. $x_{k+2} = 2x_k - x_{k+1}$, where $x_0 = 1$ and $x_1 = 2$.
c. $x_{k+2} = 2x_k + x_{k+1}$, where $x_0 = 0$ and $x_1 = 1$.
d. $x_{k+2} = 6x_k - x_{k+1}$, where $x_0 = 1$ and $x_1 = 1$.

Answer available at the end of this section.

Exercise $7.4E.2$

Solve the following linear recurrences.

a. $x_{k+3} = 6x_{k+2} - 11x_{k+1} + 6x_k$, where $x_0 = 1$, $x_1 = 0$, and $x_2 = 1$.
b. $x_{k+3} = -2x_{k+2} + x_{k+1} + 2x_k$, where $x_0 = 1$, $x_1 = 0$, and $x_2 = 1$.

[*Hint*: Use $\mathbf{v}_k = \begin{bmatrix} x_k \\ x_{k+1} \\ x_{k+2} \end{bmatrix}$.]

Answer available at the end of this section.

Exercise $7.4E.3$

In Example 3.6.1 suppose buses are also allowed to park, and let x_k denote the number of ways a row of k parking spaces can be filled with cars, trucks, and buses.

a. If trucks and buses take up 2 and 3 spaces respectively, show that $x_{k+3} = x_k + x_{k+1} + x_{k+2}$ for each k, and use this recurrence to compute x_{10}. [*Hint*: The eigenvalues are of little use.]
b. If buses take up 4 spaces, find a recurrence for the x_k and compute x_{10}.

Answer available at the end of this section.

Exercise $7.4E.4$

A man must climb a flight of k steps. He always takes one or two steps at a time. Thus he can climb 3 steps in the following ways: 1, 1, 1; 1, 2; or 2, 1. Find s_k, the number of ways he can climb the flight of k steps. [*Hint*: Fibonacci.]

Exercise $7.4E.5$

How many "words" of k letters can be made from the letters $\{a, b\}$ if there are no adjacent a's?

Answer available at the end of this section.

Exercise 7.4E.6

How many sequences of k flips of a coin are there with no *HH*?

Exercise 7.4E.7

Find x_k, the number of ways to make a stack of k poker chips if only red, blue, and gold chips are used and no two gold chips are adjacent. [*Hint*: Show that $x_{k+2} = 2x_{k+1} + 2x_k$ by considering how many stacks have a red, blue, or gold chip on top.]

Answer available at the end of this section.

Exercise 7.4E.8

A nuclear reactor contains α- and β-particles. In every second each α-particle splits into three β-particles, and each β-particle splits into an α-particle and two β-particles. If there is a single α-particle in the reactor at time $t = 0$, how many α-particles are there at $t = 20$ seconds? [*Hint*: Let x_k and y_k denote the number of α- and β-particles at time $t = k$ seconds. Find x_{k+1} and y_{k+1} in terms of x_k and y_k.]

Exercise 7.4E.9

The annual yield of wheat in a certain country has been found to equal the average of the yield in the previous two years. If the yields in 1990 and 1991 were 10 and 12 million tons respectively, find a formula for the yield k years after 1990. What is the long-term average yield?

Answer available at the end of this section.

Exercise 7.4E.10

Find the general solution to the recurrence $x_{k+1} = rx_k + c$ where r and c are constants. [*Hint*: Consider the cases $r = 1$ and $r \neq 1$ separately. If $r \neq 1$, you will need the identity $1 + r + r^2 + \cdots + r^{n-1} = \frac{1-r^n}{1-r}$ for $n \geq 1$.]

Exercise 7.4E.11

Consider the length 3 recurrence $x_{k+3} = ax_k + bx_{k+1} + cx_{k+2}$.

a. If $\mathbf{v}_k = \begin{bmatrix} x_k \\ x_{k+1} \\ x_{k+2} \end{bmatrix}$ and $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ a & b & c \end{bmatrix}$ show that $\mathbf{v}_{k+1} = A\mathbf{v}_k$.

b. If λ is any eigenvalue of A, show that $\mathbf{x} = \begin{bmatrix} 1 \\ \lambda \\ \lambda^2 \end{bmatrix}$ is a λ-eigenvector. [*Hint*: Show directly that $A\mathbf{x} = \lambda\mathbf{x}$.]

c. Generalize (a) and (b) to a recurrence $x_{k+4} = ax_k + bx_{k+1} + cx_{k+2} + dx_{k+3}$ of length 4.

Answer available at the end of this section.

Exercise 7.4E.12

Consider the recurrence $x_{k+2} = ax_{k+1} + bx_k + c$ where c may not be zero.

a. If $a+b \neq 1$ show that p can be found such that, if we set $y_k = x_k + p$, then $y_{k+2} = ay_{k+1} + by_k$. [Hence, the sequence x_k can be found provided y_k can be found by the methods of this section (or otherwise).]
b. Use (a) to solve $x_{k+2} = x_{k+1} + 6x_k + 5$ where $x_0 = 1$ and $x_1 = 1$.

Answer available at the end of this section.

Exercise $7.4E.13$

Consider the recurrence

$$x_{k+2} = ax_{k+1} + bx_k + c(k) \tag{7.4E.1}$$

where $c(k)$ is a function of k, and consider the related recurrence

$$x_{k+2} = ax_{k+1} + bx_k \tag{7.4E.2}$$

Suppose that $x_k = p_k$ is a particular solution of Equation 7.4E.1.

a. If q_k is any solution of Equation 7.4E.2, show that $q_k + p_k$ is a solution of Equation 7.4E.1.
b. Show that every solution of Equation 7.4E.1 arises as in (a) as the sum of a solution of Equation 7.4E.2 plus the particular solution p_k of Equation 7.4E.1.

Answer available at the end of this section.

Answers

Exercise $7.4E.1$

b. $x_k = \frac{1}{3}\left[4-(-2)^k\right]$
d. $x_k = \frac{1}{5}\left[2^{k+2}+(-3)^k\right]$

Exercise $7.4E.2$

b. $x_k = \frac{1}{2}\left[(-1)^k+1\right]$

Exercise $7.4E.3$

b. $x_{k+4} = x_k + x_{k+2} + x_{k+3}; x_{10} = 169$

Exercise $7.4E.5$

$\frac{1}{2\sqrt{5}}\left[3+\sqrt{5}\right]\lambda_1^k + (-3+\sqrt{5})\lambda_2^k$ where $\lambda_1 = \frac{1}{2}(1+\sqrt{5})$ and $\lambda_2 = \frac{1}{2}(1-\sqrt{5})$.

Exercise $7.4E.7$

$\frac{1}{2\sqrt{3}}\left[2+\sqrt{3}\right]\lambda_1^k + (-2+\sqrt{3})\lambda_2^k$ where $\lambda_1 = 1+\sqrt{3}$ and $\lambda_2 = 1-\sqrt{3}$.

Exercise $7.4E.9$

$\frac{34}{3} - \frac{4}{3}\left(-\frac{1}{2}\right)^k$. Long term $11\frac{1}{3}$ million tons.

Exercise $7.4E.11$

b. $A\begin{bmatrix}1\\ \lambda\\ \lambda^2\end{bmatrix} = \begin{bmatrix}\lambda\\ \lambda^2\\ a+b\lambda+c\lambda^2\end{bmatrix} = \begin{bmatrix}\lambda\\ \lambda^2\\ \lambda^3\end{bmatrix} = \lambda\begin{bmatrix}1\\ \lambda\\ \lambda^2\end{bmatrix}$

Exercise $7.4E.12$

b. $x_k = \frac{11}{10}3^k + \frac{11}{15}(-2)^k - \frac{5}{6}$

Exercise $7.4E.13$

a. $p_{k+2} + q_{k+2} = [ap_{k+1} + bp_k + c(k)] + [aq_{k+1} + bq_k] = a(p_{k+1} + q_{k+1}) + b(p_k + q_k) + c(k)$

This page titled 7.4E: An Application to Linear Recurrences Exercises is shared under a license and was authored, remixed, and/or curated by LibreTexts.

- 3.6E: An Application to Linear Recurrences Exercises by W. Keith Nicholson is licensed CC BY-NC-SA 4.0.

7.5: An Application to Systems of Differential Equations

A function f of a real variable is said to be **differentiable** if its derivative exists and, in this case, we let f' denote the derivative. If f and g are differentiable functions, a system

$$\begin{aligned} f' &= 3f + 5g \\ g' &= -f + 2g \end{aligned}$$

is called a *system of first order differential equations,* or a *differential system* for short. Solving many practical problems often comes down to finding sets of functions that satisfy such a system (often involving more than two functions). In this section we show how diagonalization can help. Of course an acquaintance with calculus is required.

The Exponential Function

The simplest differential system is the following single equation:

$$f' = af \text{ where } a \text{ is constant} \tag{7.5.1}$$

It is easily verified that $f(x) = e^{ax}$ is one solution; in fact, Equation 7.5.1 is simple enough for us to find *all* solutions. Suppose that f is any solution, so that $f'(x) = af(x)$ for all x. Consider the new function g given by $g(x) = f(x)e^{-ax}$. Then the product rule of differentiation gives

$$\begin{aligned} g'(x) &= f(x)\left[-ae^{-ax}\right] + f'(x)e^{-ax} \\ &= -af(x)e^{-ax} + \left[af(x)\right]e^{-ax} \\ &= 0 \end{aligned}$$

for all x. Hence the function $g(x)$ has zero derivative and so must be a constant, say $g(x) = c$. Thus $c = g(x) = f(x)e^{-ax}$, that is

$$f(x) = ce^{ax}$$

In other words, every solution $f(x)$ of Equation 7.5.1 is just a scalar multiple of e^{ax}. Since every such scalar multiple is easily seen to be a solution of Equation 7.5.1, we have proved

Theorem 7.5.1

The set of solutions to $f' = af$ is $\{ce^{ax} \mid c \text{ any constant}\} = \mathbb{R}e^{ax}$.

Remarkably, this result together with diagonalization enables us to solve a wide variety of differential systems.

Example 7.5.1

Assume that the number $n(t)$ of bacteria in a culture at time t has the property that the rate of change of n is proportional to n itself. If there are n_0 bacteria present when $t = 0$, find the number at time t.

Solution

Let k denote the proportionality constant. The rate of change of $n(t)$ is its time-derivative $n'(t)$, so the given relationship is $n'(t) = kn(t)$. Thus Theorem 3.7.1 shows that all solutions n are given by $n(t) = ce^{kt}$, where c is a constant. In this case, the constant c is determined by the requirement that there be n_0 bacteria present when $t = 0$. Hence $n_0 = n(0) = ce^{k0} = c$, so

$$n(t) = n_0e^{kt}$$

gives the number at time t. Of course the constant k depends on the strain of bacteria.

The condition that $n(0) = n_0$ in Example 3.7.1 is called an **initial condition** or a **boundary condition** and serves to select one solution from the available solutions.

General Differential Systems

Solving a variety of problems, particularly in science and engineering, comes down to solving a system of linear differential equations. Diagonalization enters into this as follows. The general problem is to find differentiable functions $f_1, f_2, \ldots, f_n$ that satisfy a system of equations of the form

$$\begin{array}{rcccccc} f_1' & = & a_{11}f_1 & + & a_{12}f_2 & +\cdots+ & a_{1n}f_n \\ f_2' & = & a_{21}f_1 & + & a_{22}f_2 & +\cdots+ & a_{2n}f_n \\ \vdots & & \vdots & & \vdots & & \vdots \\ f_n' & = & a_{n1}f_1 & + & a_{n2}f_2 & +\cdots+ & a_{nn}f_n \end{array}$$

where the a_{ij} are constants. This is called a **linear system of differential equations** or simply a **differential system**. The first step is to put it in matrix form. Write

$$\mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} \quad \mathbf{f}' = \begin{bmatrix} f_1' \\ f_2' \\ \vdots \\ f_n' \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Then the system can be written compactly using matrix multiplication:

$$\mathbf{f}' = A\mathbf{f}$$

Hence, given the matrix A, the problem is to find a column $\mathbf{f}$ of differentiable functions that satisfies this condition. This can be done if A is diagonalizable. Here is an example.

Example 7.5.2

Find a solution to the system

$$\begin{array}{rcl} f_1' & = & f_1 + 3f_2 \\ f_2' & = & 2f_1 + 2f_2 \end{array}$$

that satisfies $f_1(0) = 0$, $f_2(0) = 5$.

Solution

This is $\mathbf{f}' = A\mathbf{f}$, where $\mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix}$. The reader can verify that $c_A(x) = (x-4)(x+1)$, and that $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$ are eigenvectors corresponding to the eigenvalues 4 and -1, respectively. Hence the diagonalization algorithm gives $P^{-1}AP = \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix}$, where $P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & -2 \end{bmatrix}$. Now consider new functions g_1 and g_2 given by $\mathbf{f} = P\mathbf{g}$ (equivalently, $\mathbf{g} = P^{-1}\mathbf{f}$), where $\mathbf{g} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$ Then

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \quad \text{that is,} \quad \begin{array}{l} f_1 = g_1 + 3g_2 \\ f_2 = g_1 - 2g_2 \end{array}$$

Hence $f_1' = g_1' + 3g_2'$ and $f_2' = g_1' - 2g_2'$ so that

$$\mathbf{f}' = \begin{bmatrix} f_1' \\ f_2' \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} g_1' \\ g_2' \end{bmatrix} = P\mathbf{g}'$$

If this is substituted in $\mathbf{f}' = A\mathbf{f}$, the result is $P\mathbf{g}' = AP\mathbf{g}$, whence

$$\mathbf{g}' = P^{-1}AP\mathbf{g}$$

But this means that

$$\begin{bmatrix} g_1' \\ g_2' \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}, \quad \text{so} \quad \begin{array}{l} g_1' = 4g_1 \\ g_2' = -g_2 \end{array}$$

Hence Theorem 3.7.1 gives $g_1(x) = ce^{4x}$, $g_2(x) = de^{-x}$, where c and d are constants. Finally, then,

$$\begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = P \begin{bmatrix} g_1(x) \\ g_2(x) \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} ce^{4x} \\ de^{-x} \end{bmatrix} = \begin{bmatrix} ce^{4x} + 3de^{-x} \\ ce^{4x} - 2de^{-x} \end{bmatrix}$$

so the *general solution* is

$$\begin{array}{rcl} f_1(x) & = & ce^{4x} + 3de^{-x} \\ f_2(x) & = & ce^{4x} - 2de^{-x} \end{array} \qquad c \text{ and } d \text{ constants}$$

It is worth observing that this can be written in matrix form as

$$\begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = c \begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{4x} + d \begin{bmatrix} 3 \\ -2 \end{bmatrix} e^{-x}$$

That is,

$$\mathbf{f}(x) = c\mathbf{x}_1 e^{4x} + d\mathbf{x}_2 e^{-x}$$

This form of the solution works more generally, as will be shown.

Finally, the requirement that $f_1(0) = 0$ and $f_2(0) = 5$ in this example determines the constants c and d:

$$\begin{aligned} 0 &= f_1(0) = ce^0 + 3de^0 = c + 3d \\ 5 &= f_2(0) = ce^0 - 2de^0 = c - 2d \end{aligned}$$

These equations give $c = 3$ and $d = -1$, so

$$\begin{aligned} f_1(x) &= 3e^{4x} - 3e^{-x} \\ f_2(x) &= 3e^{4x} + 2e^{-x} \end{aligned}$$

satisfy all the requirements.

The technique in this example works in general.

Theorem 7.5.2

Consider a linear system

$$\mathbf{f}' = A\mathbf{f}$$

of differential equations, where A is an $n \times n$ diagonalizable matrix. Let $P^{-1}AP$ be diagonal, where P is given in terms of its columns

$$P = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$$

and $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ are eigenvectors of A. If $\mathbf{x}_i$ corresponds to the eigenvalue λ_i for each i, then every solution $\mathbf{f}$ of $\mathbf{f}' = A\mathbf{f}$ has the form

$$\mathbf{f}(x) = c_1\mathbf{x}_1e^{\lambda_1 x} + c_2\mathbf{x}_2e^{\lambda_2 x} + \cdots + c_n\mathbf{x}_ne^{\lambda_n x}$$

where $c_1, c_2, \ldots, c_n$ are arbitrary constants.

Proof

By Theorem 7.5.2, the matrix $P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$ is invertible and

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

As in Example 2.7.2, write $\mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}$ and define $\mathbf{g} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix}$ by $\mathbf{g} = P^{-1}\mathbf{f}$; equivalently, $\mathbf{f} = P\mathbf{g}$. If $P = [p_{ij}]$, this gives

$$f_i = p_{i1}g_1 + p_{i2}g_2 + \cdots + p_{in}g_n$$

Since the p_{ij} are constants, differentiation preserves this relationship:

$$f_i' = p_{i1}g_1' + p_{i2}g_2' + \cdots + p_{in}g_n'$$

so $\mathbf{f}' = P\mathbf{g}'$. Substituting this into $\mathbf{f}' = A\mathbf{f}$ gives $P\mathbf{g}' = AP\mathbf{g}$. But then left multiplication by P^{-1} gives $\mathbf{g}' = P^{-1}AP\mathbf{g}$, so the original system of equations $\mathbf{f}' = A\mathbf{f}$ for $\mathbf{f}$ becomes much simpler in terms of $\mathbf{g}$:

$$\begin{bmatrix} g_1' \\ g_2' \\ \vdots \\ g_n' \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix}$$

Hence $g_i' = \lambda_i g_i$ holds for each i, and Theorem 3.7.1 implies that the only solutions are

$$g_i(x) = c_ie^{\lambda_i x} \quad c_i \text{ some constant}$$

Then the relationship $\mathbf{f} = P\mathbf{g}$ gives the functions $f_1, f_2, \ldots, f_n$ as follows:

$$\mathbf{f}(x) = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \begin{bmatrix} c_1e^{\lambda_1 x} \\ c_2e^{\lambda_2 x} \\ \vdots \\ c_ne^{\lambda_n x} \end{bmatrix} = c_1\mathbf{x}_1e^{\lambda_1 x} + c_2\mathbf{x}_2e^{\lambda_2 x} + \cdots + c_n\mathbf{x}_ne^{\lambda_n x}$$

This is what we wanted.

□

The theorem shows that *every* solution to $\mathbf{f}' = A\mathbf{f}$ is a linear combination

$$\mathbf{f}(x) = c_1\mathbf{x}_1e^{\lambda_1 x} + c_2\mathbf{x}_2e^{\lambda_2 x} + \cdots + c_n\mathbf{x}_ne^{\lambda_n x}$$

where the coefficients c_i are arbitrary. Hence this is called the **general solution** to the system of differential equations. In most cases the solution functions $f_i(x)$ are required to satisfy boundary conditions, often of the form $f_i(a) = b_i$, where $a, b_1, \ldots, b_n$ are prescribed numbers. These conditions determine the constants c_i. The following example illustrates this and displays a situation where one eigenvalue has multiplicity greater than 1.

Example 7.5.3

Find the general solution to the system

$$\begin{array}{rcrcrcr} f_1' & = & 5f_1 & + & 8f_2 & + & 16f_3 \\ f_2' & = & 4f_1 & + & f_2 & + & 8f_3 \\ f_3' & = & -4f_1 & - & 4f_2 & - & 11f_3 \end{array}$$

Then find a solution satisfying the boundary conditions $f_1(0) = f_2(0) = f_3(0) = 1$.

Solution

The system has the form $\mathbf{f}' = A\mathbf{f}$, where $A = \begin{bmatrix} 5 & 8 & 16 \\ 4 & 1 & 8 \\ -4 & -4 & -11 \end{bmatrix}$. In this case $c_A(x) = (x+3)^2(x-1)$ and eigenvectors corresponding to the eigenvalues -3, -3, and 1 are, respectively,

$$\mathbf{x}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

Hence, by Theorem 3.7.2, the general solution is

$$\mathbf{f}(x) = c_1 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} e^{-3x} + c_2 \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} e^{-3x} + c_3 \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} e^{x}, \quad c_i \text{ constants.}$$

The boundary conditions $f_1(0) = f_2(0) = f_3(0) = 1$ determine the constants c_i.

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{f}(0) = c_1 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$
$$= \begin{bmatrix} -1 & -2 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

The solution is $c_1 = -3$, $c_2 = 5$, $c_3 = 4$, so the required specific solution is

$$\begin{array}{rcrcrcr} f_1(x) = & & -7e^{-3x} & + & 8e^{x} \\ f_2(x) = & & -3e^{-3x} & + & 4e^{x} \\ f_3(x) = & & 5e^{-3x} & - & 4e^{x} \end{array}$$

This page titled 7.5: An Application to Systems of Differential Equations is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by W. Keith Nicholson via source content that was edited to the style and standards of the LibreTexts platform.

- 3.7: An Application to Systems of Differential Equations by W. Keith Nicholson is licensed CC BY-NC-SA 4.0. Original source: https://lyryx.com/linear-algebra-applications.

7.5E: An Application to Systems of Differential Equations Exercises

Exercise $7.5E.1$

Use Theorem 3.7.1 to find the general solution to each of the following systems. Then find a specific solution satisfying the given boundary condition.

a. $\begin{aligned} f_1' &= 2f_1 + 4f_2, & f_1(0) &= 0 \\ f_2' &= 3f_1 + 3f_2, & f_2(0) &= 1 \end{aligned}$

b. $\begin{aligned} f_1' &= -f_1 + 5f_2, & f_1(0) &= 1 \\ f_2' &= f_1 + 3f_2, & f_2(0) &= -1 \end{aligned}$

c. $\begin{aligned} f_1' &= 4f_2 + 4f_3 & f_1(0) = f_2(0) = f_3(0) &= 1 \\ f_2' &= f_1 + f_2 - 2f_3 \\ f_3' &= -f_1 + f_2 + 4f_3 \end{aligned}$

d. $\begin{aligned} f_1' &= 2f_1 + f_2 + 2f_3 & f_1(0) = f_2(0) = f_3(0) &= 1 \\ f_2' &= 2f_1 + 2f_2 - 2f_3 \\ f_3' &= 3f_1 + f_2 + f_3 \end{aligned}$

Answer available at the end of this section.

Exercise $7.5E.2$

Show that the solution to $f' = af$ satisfying $f(x_0) = k$ is $f(x) = ke^{a(x-x_0)}$.

Exercise $7.5E.3$

A radioactive element decays at a rate proportional to the amount present. Suppose an initial mass of 10 g decays to 8 g in 3 hours.

a. Find the mass t hours later.
b. Find the half-life of the element—the time taken to decay to half its mass.

Answer available at the end of this section.

Exercise $7.5E.4$

The population $N(t)$ of a region at time t increases at a rate proportional to the population. If the population doubles every 5 years and is 3 million initially, find $N(t)$.

Exercise $7.5E.5$

Let A be an invertible diagonalizable $n \times n$ matrix and let $\mathbf{b}$ be an n-column of constant functions. We can solve the system $\mathbf{f}' = A\mathbf{f} + \mathbf{b}$ as follows:

a. If $\mathbf{g}$ satisfies $\mathbf{g}' = A\mathbf{g}$ (using Theorem 3.7.2), show that $\mathbf{f} = \mathbf{g} - A^{-1}\mathbf{b}$ is a solution to $\mathbf{f}' = A\mathbf{f} + \mathbf{b}$.
b. Show that every solution to $\mathbf{f}' = A\mathbf{f} + \mathbf{b}$ arises as in (a) for some solution $\mathbf{g}$ to $\mathbf{g}' = A\mathbf{g}$.

Answer available at the end of this section.

Exercise $7.5E.6$

Denote the second derivative of f by $f'' = (f')'$. Consider the second order differential equation

$$f'' - a_1 f' - a_2 f = 0, \quad a_1 \text{ and } a_2 \text{ real numbers} \tag{7.5E.1}$$

a. If f is a solution to Equation 7.5E.1 let $f_1 = f$ and $f_2 = f' - a_1 f$. Show that $\begin{cases} f_1' = a_1 f_1 + f_2 \\ f_2' = a_2 f_1 \end{cases}$, that is $\begin{bmatrix} f_1' \\ f_2' \end{bmatrix} = \begin{bmatrix} a_1 & 1 \\ a_2 & 0 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$

b. Conversely, if $\begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$ is a solution to the system in (a), show that f_1 is a solution to Equation 7.5E.1.

Answer available at the end of this section.

Exercise $7.5E.7$

Writing $f''' = (f'')'$, consider the third order differential equation

$$f''' - a_1 f'' - a_2 f' - a_3 f = 0$$

where a_1, a_2, and a_3 are real numbers. Let $f_1 = f$, $f_2 = f' - a_1 f$ and $f_3 = f'' - a_1 f' - a_2 f''$.

a. Show that $\begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}$ is a solution to the system

$$\begin{cases} f_1' = a_1 f_1 + f_2 \\ f_2' = a_2 f_1 + f_3 \\ f_3' = a_3 f_1 \end{cases},$$

that is

$$\begin{bmatrix} f_1' \\ f_2' \\ f_3' \end{bmatrix} = \begin{bmatrix} a_1 & 1 & 0 \\ a_2 & 0 & 1 \\ a_3 & 0 & 0 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}$$

b. Show further that if $\begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}$ is any solution to this system, then $f = f_1$ is a solution to Equation 7.5E.1. *Remark*. A similar construction casts every linear differential equation of order n (with constant coefficients) as an $n \times n$ linear system of first order equations. However, the matrix need not be diagonalizable, so other methods have been developed

Answers

Exercise $7.5E.1$

b. $c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{4x} + c_2 \begin{bmatrix} 5 \\ -1 \end{bmatrix} e^{-2x}; c_1 = -\frac{2}{3}, c_2 = \frac{1}{3}$

d. $c_1 \begin{bmatrix} -8 \\ 10 \\ 7 \end{bmatrix} e^{-x} + c_2 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} e^{2x} + c_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} e^{4x}; c_1 = 0, c_2 = -\frac{1}{2}, c_3 = \frac{3}{2}$

Exercise $7.5E.3$

b. The solution to (a) is $m(t) = 10\left(\frac{4}{5}\right)^{t/3}$. Hence we want t such that $10\left(\frac{4}{5}\right)^{t/3} = 5$. We solve for t by taking natural logarithms:

$$t = \frac{3\ln(\frac{1}{2})}{\ln(\frac{4}{5})} = 9.32 \text{ hours.}$$

Exercise $7.5E.5$

a. If $\mathbf{g}' = A\mathbf{g}$, put $\mathbf{f} = \mathbf{g} - A^{-1}\mathbf{b}$. Then $\mathbf{f}' = \mathbf{g}'$ and $A\mathbf{f} = A\mathbf{g} - \mathbf{b}$, so $\mathbf{f}' = \mathbf{g}' = A\mathbf{g} = A\mathbf{f} + \mathbf{b}$, as required.

Exercise $7.5E.6$

b. Assume that $f_1' = a_1 f_1 + f_2$ and $f_2' = a_2 f_1$. Differentiating gives $f_1'' = a_1 f_1' + f_2' = a_1 f_1' + a_2 f_1$, proving that f_1 satisfies Equation 7.5E.1.

This page titled 7.5E: An Application to Systems of Differential Equations Exercises is shared under a license and was authored, remixed, and/or curated by LibreTexts.

- 3.7E: An Application to Systems of Differential Equations Exercises by W. Keith Nicholson is licensed CC BY-NC-SA 4.0.

7.6: An Application to Computer Graphics

Computer graphics deals with images displayed on a computer screen, and so arises in a variety of applications, ranging from word processors, to *Star Wars* animations, to video games, to wire-frame images of an airplane. These images consist of a number of points on the screen, together with instructions on how to fill in areas bounded by lines and curves. Often curves are approximated by a set of short straight-line segments, so that the curve is specified by a series of points on the screen at the end of these segments. Matrix transformations are important here because matrix images of straight line segments are again line segments.[1] Note that a colour image requires that three images are sent, one to each of the red, green, and blue phosphorus dots on the screen, in varying intensities.

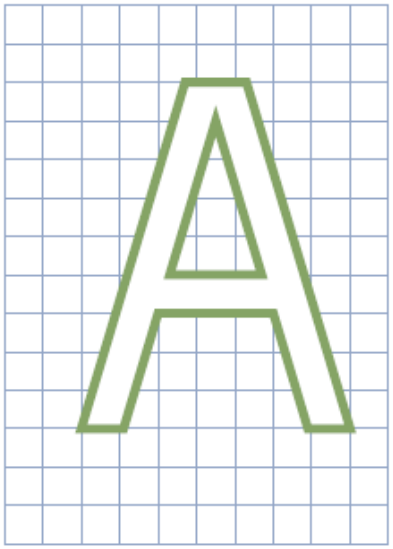


Figure 7.6.1

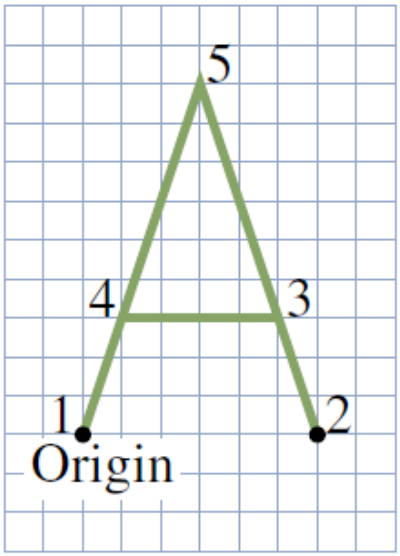


Figure 7.6.2

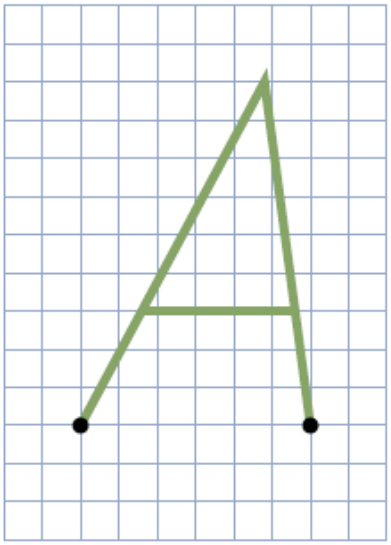


Figure 7.6.3

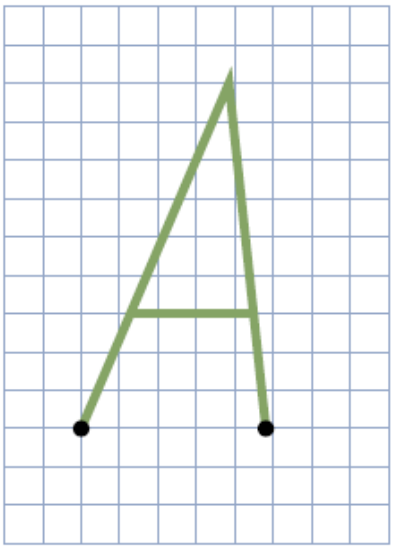


Figure 7.6.4

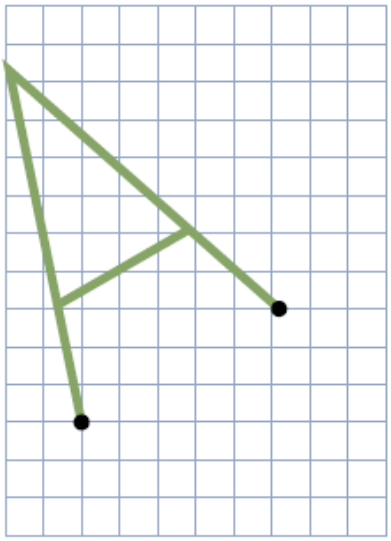


Figure 7.6.5

Consider displaying the letter A. In reality, it is depicted on the screen, as in Figure 7.6.1, by specifying the coordinates of the 11 corners and filling in the interior.

For simplicity, we will disregard the thickness of the letter, so we require only five coordinates as in Figure 7.6.2.

This simplified letter can then be stored as a data matrix

$$\begin{matrix} Vertex & 1 & 2 & 3 & 4 & 5 \end{matrix}$$
$$D = \begin{bmatrix} 0 & 6 & 5 & 1 & 3 \\ 0 & 0 & 3 & 3 & 9 \end{bmatrix}$$

where the columns are the coordinates of the vertices in order. Then if we want to transform the letter by a 2×2 matrix A, we left-multiply this data matrix by A (the effect is to multiply each column by A and so transform each vertex).

For example, we can slant the letter to the right by multiplying by an x-shear matrix $A = \begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix}$ —see Section 2.2. The result is the letter with data matrix

$$A = \begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 6 & 5 & 1 & 3 \\ 0 & 0 & 3 & 3 & 9 \end{bmatrix} = \begin{bmatrix} 0 & 6 & 5.6 & 1.6 & 4.8 \\ 0 & 0 & 3 & 3 & 9 \end{bmatrix}$$

which is shown in Figure 7.6.3.

If we want to make this slanted matrix narrower, we can now apply an x-scale matrix $B = \begin{bmatrix} 0.8 & 0 \\ 0 & 1 \end{bmatrix}$ that shrinks the x-coordinate by 0.8. The result is the composite transformation

$$BAD = \begin{bmatrix} 0.8 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 6 & 5 & 1 & 3 \\ 0 & 0 & 3 & 3 & 9 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 4.8 & 4.48 & 1.28 & 3.84 \\ 0 & 0 & 3 & 3 & 9 \end{bmatrix}$$

which is drawn in Figure 7.6.4.

On the other hand, we can rotate the letter about the origin through $\frac{\pi}{6}$ (or 30°) by multiplying by the matrix $R_{\frac{\pi}{2}} = \begin{bmatrix} \cos(\frac{\pi}{6}) & -\sin(\frac{\pi}{6}) \\ \sin(\frac{\pi}{6}) & \cos(\frac{\pi}{6}) \end{bmatrix} = \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix}$. This gives

$$R_{\frac{\pi}{2}} = \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix} \begin{bmatrix} 0 & 6 & 5 & 1 & 3 \\ 0 & 0 & 3 & 3 & 9 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 5.196 & 2.83 & -0.634 & -1.902 \\ 0 & 3 & 5.098 & 3.098 & 9.294 \end{bmatrix}$$

and is plotted in Figure 7.6.5.

This poses a problem: How do we rotate at a point other than the origin? It turns out that we can do this when we have solved another more basic problem. It is clearly important to be able to translate a screen image by a fixed vector $\mathbf{w}$, that is apply the transformation $T_w : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T_w(\mathbf{v}) = \mathbf{v} + \mathbf{w}$ for all $\mathbf{v}$ in $\mathbb{R}^2$. The problem is that these translations are not matrix transformations $\mathbb{R}^2 \to \mathbb{R}^2$ because they do not carry $\mathbf{0}$ to $\mathbf{0}$ (unless $\mathbf{w} = \mathbf{0}$). However, there is a clever way around this.

The idea is to represent a point $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ as a 3×1 column $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$, called the **homogeneous coordinates** of $\mathbf{v}$. Then translation by $\mathbf{w} = \begin{bmatrix} p \\ q \end{bmatrix}$ can be achieved by multiplying by a 3×3 matrix:

$$\begin{bmatrix} 1 & 0 & p \\ 0 & 1 & q \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x+p \\ y+q \\ 1 \end{bmatrix} = \begin{bmatrix} T_{\mathbf{w}}(\mathbf{v}) \\ 1 \end{bmatrix}$$

Thus, by using homogeneous coordinates we can implement the translation T_w in the top two coordinates. On the other hand, the matrix transformation induced by $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is also given by a 3×3 matrix:

$$\begin{bmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} ax+by \\ cx+dy \\ 1 \end{bmatrix} = \begin{bmatrix} A\mathbf{v} \\ 1 \end{bmatrix}$$

So everything can be accomplished at the expense of using 3×3 matrices and homogeneous coordinates.

Example 7.6.1

Rotate the letter A in Figure 7.6.2 through $\frac{\pi}{6}$ about the point $\begin{bmatrix} 4 \\ 5 \end{bmatrix}$.

Solution

Using homogeneous coordinates for the vertices of the letter results in a data matrix with three rows:

$$K_d = \begin{bmatrix} 0 & 6 & 5 & 1 & 3 \\ 0 & 0 & 3 & 3 & 9 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

If we write $\mathbf{w} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$, the idea is to use a composite of transformations: First translate the letter by $-\mathbf{w}$ so that the point $\mathbf{w}$ moves to the origin, then rotate this translated letter, and then translate it by $\mathbf{w}$ back to its original position. The matrix arithmetic is as follows (remember the order of composition!):

$$\begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.866 & -0.5 & 0 \\ 0.5 & 0.866 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & -5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 6 & 5 & 1 & 3 \\ 0 & 0 & 3 & 3 & 9 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 3.036 & 8.232 & 5.866 & 2.402 & 1.134 \\ -1.33 & 1.67 & 3.768 & 1.768 & 7.964 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

This is plotted in Figure 7.6.6.

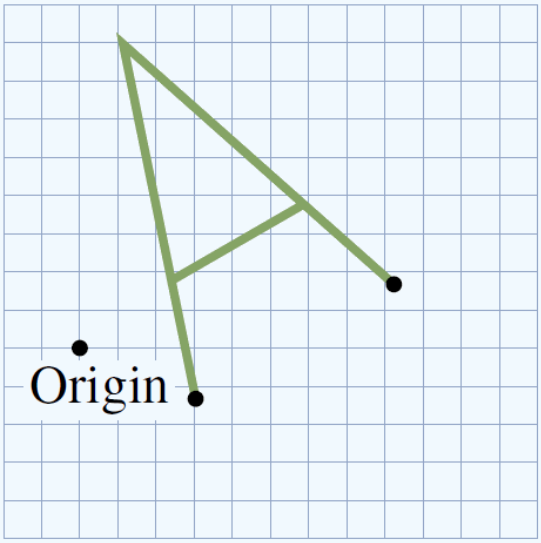


Figure 7.6.6

This discussion merely touches the surface of computer graphics, and the reader is referred to specialized books on the subject. Realistic graphic rendering requires an enormous number of matrix calculations. In fact, matrix multiplication algorithms are now embedded in microchip circuits, and can perform over 100 million matrix multiplications per second. This is particularly important in the field of three-dimensional graphics where the homogeneous coordinates have four components and 4×4 matrices are required.

---

1. If $\mathbf{v}_0$ and $\mathbf{v}_1$ are vectors, the vector from $\mathbf{v}_0$ to $\mathbf{v}_1$ is $\mathbf{d} = \mathbf{v}_1 - \mathbf{v}_0$. So a vector $\mathbf{v}$ lies on the line segment between $\mathbf{v}_0$ and $\mathbf{v}_1$ if and only if $\mathbf{v} = \mathbf{v}_0 + t\mathbf{d}$ for some number t in the range $0 \leq t \leq 1$. Thus the image of this segment is the set of vectors $A\mathbf{v} = A\mathbf{v}_0 + tA\mathbf{d}$ with $0 \leq t \leq 1$, that is the image is the segment between $A\mathbf{v}_0$ and $A\mathbf{v}_1$.

This page titled 7.6: An Application to Computer Graphics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by W. Keith Nicholson via source content that was edited to the style and standards of the LibreTexts platform.

- 4.5: An Application to Computer Graphics by W. Keith Nicholson is licensed CC BY-NC-SA 4.0. Original source: https://lyryx.com/linear-algebra-applications.

7.6E: An Application to Computer Graphics Exercises

Exercise $7.6E.1$

Consider the letter A described in Figure [fig:013290]. Find the data matrix for the letter obtained by:

a. Rotating the letter through $\frac{\pi}{4}$ about the origin.
b. Rotating the letter through $\frac{\pi}{4}$ about the point $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

Answer available at the end of this section.

Exercise $7.6E.2$

Find the matrix for turning the letter A in Figure [fig:013290] upside-down in place.

Exercise $7.6E.3$

Find the 3×3 matrix for reflecting in the line $y = mx + b$. Use $\begin{bmatrix} 1 \\ m \end{bmatrix}$ as direction vector for the line.

Exercise $7.6E.4$

Find the 3×3 matrix for rotating through the angle θ about the point $P(a, b)$.

Exercise $7.6E.5$

Find the reflection of the point P in the line $y = 1 + 2x$ in $\mathbb{R}^2$ if:

b. $P = P(1, 1)$
c. $P = P(1, 4)$
d. What about $P = P(1, 3)$? Explain. [*Hint*: Example 4.5.1 and Section 4.4.]

Answer available at the end of this section.

Answers

Exercise $7.6E.1$

b. $\frac{1}{2}\begin{bmatrix} \sqrt{2}+2 & 7\sqrt{2}+2 & 3\sqrt{2}+2 & -\sqrt{2}+2 & -5\sqrt{2}+2 \\ -3\sqrt{2}+4 & 3\sqrt{2}+4 & 5\sqrt{2}+4 & \sqrt{2}+4 & 9\sqrt{2}+4 \\ 2 & 2 & 2 & 2 & 2 \end{bmatrix}$

Exercise $7.6E.5$

b. $P(\frac{9}{5}, \frac{18}{5})$

This page titled 7.6E: An Application to Computer Graphics Exercises is shared under a license and was authored, remixed, and/or curated by LibreTexts.

- 4.5E: An Application to Computer Graphics Exercises by W. Keith Nicholson is licensed CC BY-NC-SA 4.0.

7.7: An Application to Input-Output Economic Models

In 1973 Wassily Leontief was awarded the Nobel prize in economics for his work on mathematical models.[2] Roughly speaking, an economic system in this model consists of several industries, each of which produces a product and each of which uses some of the production of the other industries. The following example is typical.

Example 7.7.1

A primitive society has three basic needs: food, shelter, and clothing. There are thus three industries in the society—the farming, housing, and garment industries—that produce these commodities. Each of these industries consumes a certain proportion of the total output of each commodity according to the following table.

Consumption	Output (Farming)	Output (Housing)	Output (Garment)
Farming	0.4	0.2	0.3
Housing	0.2	0.6	0.4
Garment	0.4	0.2	0.3

Find the annual prices that each industry must charge for its income to equal its expenditures.

Solution

Let p_1, p_2, and p_3 be the prices charged per year by the farming, housing, and garment industries, respectively, for their total output. To see how these prices are determined, consider the farming industry. It receives p_1 for its production in any year. But it *consumes* products from all these industries in the following amounts (from row 1 of the table): 40% of the food, 20% of the housing, and 30% of the clothing. Hence, the expenditures of the farming industry are $0.4p_1 + 0.2p_2 + 0.3p_3$, so

$$0.4p_1 + 0.2p_2 + 0.3p_3 = p_1$$

A similar analysis of the other two industries leads to the following system of equations.

$$\begin{aligned} 0.4p_1 + 0.2p_2 + 0.3p_3 &= p_1 \\ 0.2p_1 + 0.6p_2 + 0.4p_3 &= p_2 \\ 0.4p_1 + 0.2p_2 + 0.3p_3 &= p_3 \end{aligned}$$

This has the matrix form $E\mathbf{p} = \mathbf{p}$, where

$$E = \begin{bmatrix} 0.4 & 0.2 & 0.3 \\ 0.2 & 0.6 & 0.4 \\ 0.4 & 0.2 & 0.3 \end{bmatrix} \quad \text{and} \quad \mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$$

The equations can be written as the homogeneous system

$$(I - E)\mathbf{p} = \mathbf{0}$$

where I is the 3×3 identity matrix, and the solutions are

$$\mathbf{p} = \begin{bmatrix} 2t \\ 3t \\ 2t \end{bmatrix}$$

where t is a parameter. Thus, the pricing must be such that the total output of the farming industry has the same value as the total output of the garment industry, whereas the total value of the housing industry must be $\frac{3}{2}$ as much.

In general, suppose an economy has n industries, each of which uses some (possibly none) of the production of every industry. We assume first that the economy is **closed** (that is, no product is exported or imported) and that all product is used. Given two industries i and j, let e_{ij} denote the proportion of the total annual output of industry j that is consumed by industry i. Then $E = [e_{ij}]$ is called the **input-output** matrix for the economy. Clearly,

$$0 \leq e_{ij} \leq 1 \quad \text{for all } i \text{ and } j \tag{7.7.1}$$

Moreover, all the output from industry j is used by *some* industry (the model is closed), so

$$e_{1j} + e_{2j} + \cdots + e_{ij} = 1 \quad \text{for each } j \tag{7.7.2}$$

This condition asserts that each column of E sums to 1. Matrices satisfying conditions (7.7.1) and (7.7.2) are called **stochastic matrices**.

As in Example 2.8.1, let p_i denote the price of the total annual production of industry i. Then p_i is the annual revenue of industry i. On the other hand, industry i spends $e_{i1}p_1 + e_{i2}p_2 + \cdots + e_{in}p_n$ annually for the product it uses ($e_{ij}p_j$ is the cost for product from industry j). The closed economic system is said to be in **equilibrium** if the annual expenditure equals the annual revenue for each industry—that is, if

$$e_{1j}p_1 + e_{2j}p_2 + \cdots + e_{ij}p_n = p_i \quad \text{for each } i = 1, 2, \ldots, n$$

If we write $\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}$, these equations can be written as the matrix equation

$$E\mathbf{p} = \mathbf{p}$$

This is called the **equilibrium condition**, and the solutions $\mathbf{p}$ are called **equilibrium price structures**. The equilibrium condition can be written as

$$(I - E)\mathbf{p} = \mathbf{0}$$

which is a system of homogeneous equations for $\mathbf{p}$. Moreover, there is always a nontrivial solution $\mathbf{p}$. Indeed, the column sums of $I - E$ are all 0 (because E is stochastic), so the row-echelon form of $I - E$ has a row of zeros. In fact, more is true:

Theorem 7.7.1

Let E be any $n \times n$ stochastic matrix. Then there is a nonzero $n \times 1$ vector $\mathbf{p}$ with nonnegative entries such that $E\mathbf{p} = \mathbf{p}$. If all the entries of E are positive, the matrix $\mathbf{p}$ can be chosen with all entries positive.

Theorem 2.8.1 guarantees the existence of an equilibrium price structure for any closed input-output system of the type discussed here. The proof is beyond the scope of this book.[3]

Example 7.7.2

Find the equilibrium price structures for four industries if the input-output matrix is

$$E = \begin{bmatrix} 0.6 & 0.2 & 0.1 & 0.1 \\ 0.3 & 0.4 & 0.2 & 0 \\ 0.1 & 0.3 & 0.5 & 0.2 \\ 0 & 0.1 & 0.2 & 0.7 \end{bmatrix}$$

Find the prices if the total value of business is \$1000.

Solution

If $\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix}$ is the equilibrium price structure, then the equilibrium condition reads $E\mathbf{p} = \mathbf{p}$. When we write this as $(I - E)\mathbf{p} = \mathbf{0}$, the methods of Chapter 1 yield the following family of solutions:

$$\mathbf{p} = \begin{bmatrix} 44t \\ 39t \\ 51t \\ 47t \end{bmatrix}$$

where t is a parameter. If we insist that $p_1 + p_2 + p_3 + p_4 = 1000$, then $t = 5.525$. Hence

$$\mathbf{p} = \begin{bmatrix} 243.09 \\ 215.47 \\ 281.76 \\ 259.67 \end{bmatrix}$$

to five figures.

The Open Model

We now assume that there is a demand for products in the **open sector** of the economy, which is the part of the economy other than the producing industries (for example, consumers). Let d_i denote the total value of the demand for product i in the open sector. If p_i and e_{ij} are as before, the value of the annual demand for product i by the producing industries themselves is $e_{i1}p_1 + e_{i2}p_2 + \cdots + e_{in}p_n$, so the total annual revenue p_i of industry i breaks down as follows:

$$p_i = (e_{i1}p_1 + e_{i2}p_2 + \cdots + e_{in}p_n) + d_i \quad \text{for each } i = 1, 2, \ldots, n$$

The column $\mathbf{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix}$ is called the **demand matrix**, and this gives a matrix equation

$$\mathbf{p} = E\mathbf{p} + \mathbf{d}$$

or

$$(I - E)\mathbf{p} = \mathbf{d} \tag{7.7.3}$$

This is a system of linear equations for $\mathbf{p}$, and we ask for a solution $\mathbf{p}$ with every entry nonnegative. Note that every entry of E is between 0 and 1, but the column sums of E need not equal 1 as in the closed model.

Before proceeding, it is convenient to introduce a useful notation. If $A = [a_{ij}]$ and $B = [b_{ij}]$ are matrices of the same size, we write $A > B$ if $a_{ij} > b_{ij}$ for all i and j, and we write $A \geq B$ if $a_{ij} \geq b_{ij}$ for all i and j. Thus $P \geq 0$ means that every entry of P is nonnegative. Note that $A \geq 0$ and $B \geq 0$ implies that $AB \geq 0$.

Now, given a demand matrix $\mathbf{d} \geq \mathbf{0}$, we look for a production matrix $\mathbf{p} \geq \mathbf{0}$ satisfying equation (7.7.3). This certainly exists if $I - E$ is invertible and $(I - E)^{-1} \geq 0$. On the other hand, the fact that $\mathbf{d} \geq \mathbf{0}$ means any solution $\mathbf{p}$ to equation (7.7.3) satisfies $\mathbf{p} \geq E\mathbf{p}$. Hence, the following theorem is not too surprising.

Theorem 7.7.1

Let $E \geq 0$ be a square matrix. Then $I - E$ is invertible and $(I - E)^{-1} \geq 0$ if and only if there exists a column $\mathbf{p} > \mathbf{0}$ such that $\mathbf{p} > E\mathbf{p}$.

Heuristic Proof

If $(I-E)^{-1} \geq 0$, the existence of $\mathbf{p} > \mathbf{0}$ with $\mathbf{p} > E\mathbf{p}$ is left as Exercise 2.8.11. Conversely, suppose such a column $\mathbf{p}$ exists. Observe that

$$(I-E)(I+E+E^2+\cdots+E^{k-1}) = I-E^k$$

holds for all $k \geq 2$. If we can show that every entry of E^k approaches 0 as k becomes large then, intuitively, the infinite matrix sum

$$U = I+E+E^2+\cdots$$

exists and $(I-E)U = I$. Since $U \geq 0$, this does it. To show that E^k approaches 0, it suffices to show that $EP < \mu P$ for some number μ with $0 < \mu < 1$ (then $E^kP < \mu^k P$ for all $k \geq 1$ by induction). The existence of μ is left as Exercise 2.8.12.

□

The condition $\mathbf{p} > E\mathbf{p}$ in Theorem 2.8.2 has a simple economic interpretation. If $\mathbf{p}$ is a production matrix, entry i of $E\mathbf{p}$ is the total value of all product used by industry i in a year. Hence, the condition $\mathbf{p} > E\mathbf{p}$ means that, for each i, the value of product produced by industry i exceeds the value of the product it uses. In other words, each industry runs at a profit.

Example 7.7.3

If $E = \begin{bmatrix} 0.6 & 0.2 & 0.3 \\ 0.1 & 0.4 & 0.2 \\ 0.2 & 0.5 & 0.1 \end{bmatrix}$, show that $I-E$ is invertible and $(I-E)^{-1} \geq 0$.

Solution

Use $\mathbf{p} = (3,2,2)^T$ in Theorem 2.8.2.

If $\mathbf{p}_0 = (1,1,1)^T$, the entries of $E\mathbf{p}_0$ are the row sums of E. Hence $\mathbf{p}_0 > E\mathbf{p}_0$ holds if the row sums of E are all less than 1. This proves the first of the following useful facts (the second is Exercise 2.8.10).

Corollary 7.7.1

Let $E \geq 0$ be a square matrix. In each case, $I-E$ is invertible and $(I-E)^{-1} \geq 0$:

1. All row sums of E are less than 1.
2. All column sums of E are less than 1.

---

1. The applications in this section and the next are independent and may be taken in any order.
2. See W. W. Leontief, "The world economy of the year 2000," *Scientific American,* Sept. 1980.
3. The interested reader is referred to P. Lancaster's *Theory of Matrices* (New York: Academic Press, 1969) or to E. Seneta's *Non-negative Matrices* (New York: Wiley, 1973).

This page titled 7.7: An Application to Input-Output Economic Models is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by W. Keith Nicholson via source content that was edited to the style and standards of the LibreTexts platform.

- 2.8: An Application to Input-Output Economic Models by W. Keith Nicholson is licensed CC BY-NC-SA 4.0. Original source: https://lyryx.com/linear-algebra-applications.

7.7E: An Application to Input-Output Economic Models Exercises

Exercise $7.7E.1$

Find the possible equilibrium price structures when the input-output matrices are:

a. $\begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.6 & 0.2 & 0.3 \\ 0.3 & 0.6 & 0.4 \end{bmatrix}$

b. $\begin{bmatrix} 0.5 & 0 & 0.5 \\ 0.1 & 0.9 & 0.2 \\ 0.4 & 0.1 & 0.3 \end{bmatrix}$

c. $\begin{bmatrix} 0.3 & 0.1 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.1 & 0 \\ 0.3 & 0.3 & 0.2 & 0.3 \\ 0.2 & 0.3 & 0.6 & 0.5 \end{bmatrix}$

d. $\begin{bmatrix} 0.5 & 0 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0 & 0.1 \\ 0.1 & 0.2 & 0.8 & 0.2 \\ 0.2 & 0.1 & 0.1 & 0.6 \end{bmatrix}$

Answer available at the end of this section.

Exercise $7.7E.2$

Three industries A, B, and C are such that all the output of A is used by B, all the output of B is used by C, and all the output of C is used by A. Find the possible equilibrium price structures.

Answer available at the end of this section.

Exercise $7.7E.3$

Find the possible equilibrium price structures for three industries where the input-output matrix is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$. Discuss why there are two parameters here.

Exercise $7.7E.4$

Prove Theorem 2.8.1 for a 2×2 stochastic matrix E by first writing it in the form $E = \begin{bmatrix} a & b \\ 1-a & 1-b \end{bmatrix}$, where $0 \leq a \leq 1$ and $0 \leq b \leq 1$.

Answer available at the end of this section.

Exercise 7.7E.5

If E is an $n \times n$ stochastic matrix and $\mathbf{c}$ is an $n \times 1$ matrix, show that the sum of the entries of $\mathbf{c}$ equals the sum of the entries of the $n \times 1$ matrix $E\mathbf{c}$.

Exercise 7.7E.6

Let $W = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$. Let E and F denote $n \times n$ matrices with nonnegative entries.

a. Show that E is a stochastic matrix if and only if $WE = W$.
b. Use part (a.) to deduce that, if E and F are both stochastic matrices, then EF is also stochastic.

Exercise 7.7E.7

Find a 2×2 matrix E with entries between 0 and 1 such that:

a. $I - E$ has no inverse.
b. $I - E$ has an inverse but not all entries of $(I - E)^{-1}$ are nonnegative.

Answer available at the end of this section.

Exercise 7.7E.8

If E is a 2×2 matrix with entries between 0 and 1, show that $I - E$ is invertible and $(I - E)^{-1} \geq 0$ if and only if $tr\ E < 1 + \det E$. Here, if $E = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $tr\ E = a + d$ and $\det E = ad - bc$.

Answer available at the end of this section.

Exercise 7.7E.9

In each case show that $I - E$ is invertible and $(I - E)^{-1} \geq 0$.

b. $\begin{bmatrix} 0.6 & 0.5 & 0.1 \\ 0.1 & 0.3 & 0.3 \\ 0.2 & 0.1 & 0.4 \end{bmatrix}$

c. $\begin{bmatrix} 0.7 & 0.1 & 0.3 \\ 0.2 & 0.5 & 0.2 \\ 0.1 & 0.1 & 0.4 \end{bmatrix}$

d. $\begin{bmatrix} 0.6 & 0.2 & 0.1 \\ 0.3 & 0.4 & 0.2 \\ 0.2 & 0.5 & 0.1 \end{bmatrix}$

e. $\begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.3 & 0.1 & 0.2 \\ 0.3 & 0.3 & 0.2 \end{bmatrix}$

Answer available at the end of this section.

Exercise $7.7E.10$

Prove that (1) implies (2) in the Corollary to Theorem 2.8.2.

Exercise $7.7E.11$

If $(I-E)^{-1} \geq 0$, find $\mathbf{p} > 0$ such that $\mathbf{p} > E\mathbf{p}$.

Exercise $7.7E.12$

If $E\mathbf{p} < \mathbf{p}$ where $E \geq 0$ and $\mathbf{p} > 0$, find a number μ such that $E\mathbf{p} < \mu\mathbf{p}$ and $0 < \mu < 1$.

[*Hint*: If $E\mathbf{p} = (q_1, \ldots, q_n)^T$ and $\mathbf{p} = (p_1, \ldots, p_n)^T$, take any number μ where $\max\left\{\frac{q_1}{p_1}, \ldots, \frac{q_n}{p_n}\right\} < \mu < 1$.]

Answers

Exercise $7.7E.1$

b. $\begin{bmatrix} t \\ 3t \\ t \end{bmatrix}$

d. $\begin{bmatrix} 14t \\ 17t \\ 47t \\ 23t \end{bmatrix}$

Exercise $7.7E.2$

$\begin{bmatrix} t \\ t \\ t \end{bmatrix}$

Exercise $7.7E.4$

$P = \begin{bmatrix} bt \\ (1-a)t \end{bmatrix}$ is nonzero (for some t) unless $b = 0$ and $a = 1$. In that case, $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is a solution. If the entries of E are positive, then $P = \begin{bmatrix} b \\ 1-a \end{bmatrix}$ has positive entries.

Exercise $7.7E.7$

b. $\begin{bmatrix} 0.4 & 0.8 \\ 0.7 & 0.2 \end{bmatrix}$

Exercise $7.7E.8$

If $E = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $I - E = \begin{bmatrix} 1-a & -b \\ -c & 1-d \end{bmatrix}$, so $\det(I-E) = (1-a)(1-d) - bc = 1 - tr\ E + \det E$. If $\det(I-E) \neq 0$, then $(I-E)^{-1} = \frac{1}{\det(I-E)} \begin{bmatrix} 1-d & b \\ c & 1-a \end{bmatrix}$, so $(I-E)^{-1} \geq 0$ if $\det(I-E) > 0$, that is, $tr\ E < 1 + \det E$. The converse is now clear.

Exercise $7.7E.9$

b. Use $\mathbf{p} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$ in Theorem 2.8.2.

d. $\mathbf{p} = \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix}$ in Theorem 2.8.2.

This page titled 7.7E: An Application to Input-Output Economic Models Exercises is shared under a license and was authored, remixed, and/or curated by LibreTexts.

- 2.8E: An Application to Input-Output Economic Models Exercises by W. Keith Nicholson is licensed CC BY-NC-SA 4.0.

7.8: The Complex Number System

Focus Questions

The following questions are meant to guide our study of the material in this section. After studying this section, we should understand the concepts motivated by these questions and be able to write precise, coherent answers to these questions.

- What is a complex number?
- What does it mean for two complex numbers to be equal?
- How do we add two complex numbers together?
- How do we multiply two complex numbers together?
- What is the conjugate of a complex number?
- What is the modulus of a complex number?
- How are the conjugate and modulus of a complex number related?
- How do we divide one complex number by another?

The quadratic formula $x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ allows us to find solutions to the quadratic equation $ax^2 + bx + c = 0$. For example, the solutions to the equation $x^2 + x + 1 = 0$ are

$$x = \frac{-1 \pm \sqrt{1-4}}{2} = \frac{-1 \pm \sqrt{-3}}{2}.$$

A problem arises immediately with this solution since there is no real number t with the property that $t^2 = -3$ or $t = \sqrt{-3}$. To make sense of solutions like this we introduce *complex numbers*. Although complex numbers arise naturally when solving quadratic equations, their introduction into mathematics came about from the problem of solving cubic equations.

If we use the quadratic formula to solve an equation such as $x^2 + x + 1 = 0$,

we obtain the solutions $x = \dfrac{-1 + \sqrt{-3}}{2}$ and $x = \dfrac{-1 - \sqrt{-3}}{2}$. These numbers are complex numbers and we have a special form for writing these numbers. We write them in a way that that isolates the square root of -1. To illustrate, the number

$$\frac{-1 + \sqrt{-3}}{2}$$

can be written as follows:

$$\frac{-1 + \sqrt{-3}}{2} = -\frac{1}{2} + \frac{\sqrt{-3}}{2} = -\frac{1}{2} + \frac{\sqrt{3}\sqrt{-1}}{2} = -\frac{1}{2} + \frac{\sqrt{3}}{2}\sqrt{-1}$$

Since there is no real number t satisfying $t^2 = -1$, the number $\sqrt{-1}$ is not a real number. We call $\sqrt{-1}$ an *imaginary number* and give it a special label i. Thus, $i = \sqrt{-1}$ or $i^2 = -1$. With this in mind we can write

$$\frac{-1 + \sqrt{-3}}{2} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$$

and every complex number has this special form.

Definition: Complex Numbers

A **complex number** is an object of the form

$$a + bi \tag{7.8.1}$$

where a and b are real numbers and $i^2 = -1$.

The form $a + bi$, where a and b are real numbers is called the **standard form** for a complex number. When we have a complex number of the form $z = a + bi$, the number a is called the **real part** of the complex number z and the number b is called the **imaginary part** of z. Since i is not a real number, two complex numbers $a + bi$ and $c + di$ are equal if and only if $a = c$ and $b = d$.

There is an arithmetic of complex numbers that is determined by an addition and multiplication of complex numbers. Adding and subtracting complex numbers is natural:

$$(a + bi) + (c + di) = (a + c) + (b + d)i \tag{7.8.2}$$

$$(a + bi) - (c + di) = (a - c) + (b - d)i \tag{7.8.3}$$

That is, to add (or subtract) two complex numbers we add (subtract)their real parts and add (subtract) their imaginary parts. Multiplication is also done in a natural way – to multiply two complex numbers, we simply expand the product as usual and exploit the fact that $i^2 = -1$. So the product of two complex number is

$$\begin{aligned}(a + bi) + (c + di) &= ac + (ad)i + (bc)i + (bd)i^2 \\ &= (ac - bd) + (ad + bc)i\end{aligned}$$

Complex Number Properties

It can be shown that the complex numbers satisfy many useful and familiar properties, which are similar to properties of the real numbers. If u, w, and z, are complex numbers, then

1. $w + z = z + w$
2. $u + (w + z) = (u + w) + z$
3. The complex number $0 = 0 + 0i$ is an additive identity, that is $z + 0 = z$.
4. If $z = a + bi$, then the additive inverse of z is $-z = (-a) + (-b)i$. That is, $z + (-z) = 0$.
5. $wz = zw$
6. $u(wz) = (uw)z$
7. $u(w + z) = uw + uz$
8. If $wz = 0$, then $w = 0$ or $z = 0$.

We will use these properties as needed. For example, to write the complex product $(1 + i)i$ in the form $a + bi$ with a and b real numbers, we distribute multiplication over addition and use the fact that $i^2 = -1$ to see that

$$(1 + i)i = i + i^2 = i + (-1) = (-1) + i. \tag{7.8.4}$$

For another example, if $w = 2 + i$ and $z = 3 - 2i$, we can use these properties to write wz in the standard $a + bi$ form as follows:

$$wz = (2 + i)z = 2z + iz = 2(3 - 2i) + i(3 - 2i) = (6 - 4i) + (3i - 2i^2) = 6 - 4i + 3i - 2(-1) = 8 - i \tag{7.8.5}$$

Exercise 7.8.1A

Write each of the sums or products as a complex number in standard form.

a. $(2 + 3i) + (7 - 4i)$
b. $(4 - 2i)(3 + i)$
c. $(2 + i)i - (3 + 4i)$

Answer available at the end of this section.

Exercise 7.8.1B

Use the quadratic formula to write the two solutions to the quadratic equation $x^2 - x + 2 = 0$ as complex numbers of the form $r + si$ and $u + vi$ for some real numbers r, s, u, and v.

(**Hint**: Remember: $i = \sqrt{-1}$. So we can rewrite something like $\sqrt{-4}$ as $\sqrt{-4} = \sqrt{4}\sqrt{-1} = 2i$.)

Answer available at the end of this section.

Division of Complex Numbers

We can add, subtract, and multiply complex numbers, so it is natural to ask if we can divide complex numbers. We illustrate with an example.

Example 7.8.2: Dividing by a Complex Number

Write the quotient $\dfrac{2+i}{3+i}$ as a complex number in the form $a + bi$.

Solution

This problem is rationalizing a denominator since $i = \sqrt{-1}$. So in this case we need to "remove" the imaginary part from the denominator. Recall that the product of a complex number with its conjugate is a real number, so if we multiply the numerator and denominator of $\dfrac{2+i}{3+i}$ by the complex conjugate of the denominator, we can rewrite the denominator as a real number. The steps are as follows. Multiplying the numerator and denominator by the conjugate $3 - i$ or $3 + i$ gives us

$$\frac{2+i}{3+i} = \left(\frac{2+i}{3+i}\right)\left(\frac{3-i}{3-i}\right) = \frac{(2+i)(3-i)}{(3+i)(3-i)} = \frac{(6-i^2)+(-2+3)i}{9-i^2} = \frac{7+i}{10}$$

Now we can write the final result in standard form as

$$\frac{7+i}{10} = \frac{7}{10} + \frac{1}{10}i.$$

Example 7.8.2 illustrates the general process for dividing one complex number by another. In general, we can write the quotient $\dfrac{a+bi}{c+di}$ in the form $r + si$ by multiplying numerator and denominator of our fraction by the conjugate $c - di$ of $c + di$ to see that

$$\frac{a+bi}{c+di} = \left(\frac{a+bi}{c+di}\right)\left(\frac{c-di}{c-di}\right) = \frac{(ac+bd)+(bc-ad)i}{c^2+d^2} = \frac{ac+bd}{c^2+d^2} + \frac{bc-ad}{c^2+d^2}i \tag{7.8.6}$$

Therefore, we have the formula for the quotient of two complex numbers.

Definition: Quotient of Complex Numbers

The **quotient** $\dfrac{a+bi}{c+di}$ of the complex numbers $a + bi$ and $c + di$ is the complex number

$$\frac{a+bi}{c+di} = \frac{ac+bd}{c^2+d^2} + \frac{bc-ad}{c^2+d^2}i \tag{7.8.7}$$

provided $c + di \neq 0$.

Exercise 7.8.3

Let $z = 3 + 4i$ and $w = 5 - i$.

a. Write $\frac{w}{z} = \frac{5-i}{3+4i}$ as a complex number in the form $r + si$ where r and s are some real numbers. Check the result by multiplying the quotient by $3 + 4i$. Is this product equal to $5 - i$?
b. Find the solution to the equation $(3 + 4i)x = 5 - i$ as a complex number in the form $x = u + vi$ where u and v are some real numbers.

Answer available at the end of this section.

Geometric Representations of Complex Numbers

Each ordered pair (a, b) of real numbers determines:

- A point in the coordinate plane with coordinates (a, b).
- A complex number $a + bi$
- A vector $a\mathbf{i} + b\mathbf{j} = (a, b)$

This means that we can geometrically represent the complex number $a + bi$ with a vector in standard position with terminal point (a, b). Therefore, we can draw pictures of complex numbers in the plane. When we do this, the horizontal axis is called **the real axis**, and the vertical axis is called **the imaginary axis**. In addition, the coordinate plane is then referred to as **the complex plane**. That is, if $z = a + bi$, we can think of z as a directed line segment from the origin to the point (a, b), where the terminal point of the segment is a units from the imaginary axis and b units from the real axis. For example, the complex numbers $3 + 4i$ and $-8 + 3i$ are shown in Figure 5.1.

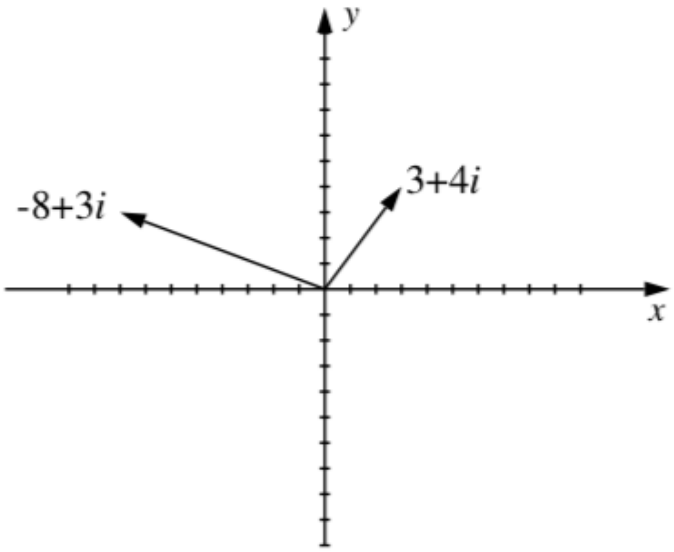


Figure 7.8.1: Two complex numbers.

In addition, the sum of two complex numbers can be represented geometrically using the vector forms of the complex numbers. Draw the parallelogram defined by $w = a + bi$ and $z = c + di$. The sum of w and z is the complex number represented by the vector from the origin to the vertex on the parallelogram opposite the origin as illustrated with the vectors $w = 3 + 4i$ and $z = -8 + 3i$ in Figure 7.8.2.

Exercise 7.8.4

Let $w = 2 + 3i$ and $z = -1 + 5i$.

1. Write the complex sum $w + z$ in standard form.
2. Draw a picture to illustrate the sum using vectors to represent w and z.

Answer available at the end of this section.

We now extend our use of the representation of a complex number as a vector in standard position to include the notion of the length of a vector. Recall from Section 3.6 that the length of a vector $\mathbf{v} = a\mathbf{i} + b\mathbf{j}$ is $|\mathbf{v}| = \sqrt{a^2 + b^2}$.

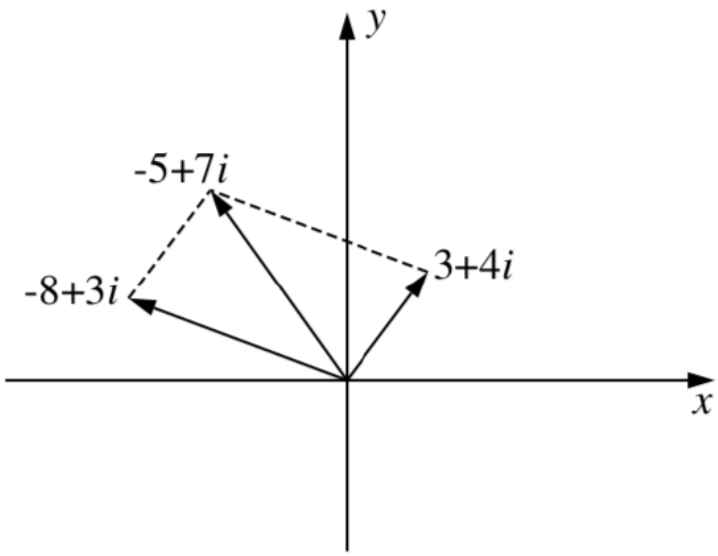


Figure 7.8.2: The Sum of Two Complex Numbers.

When we use this idea with complex numbers, we call it the *norm* or *modulus* of the complex number.

Definition: Norm

The **norm** (or **modulus**) of the complex number $z = a + bi$ is the distance from the origin to the point (a, b) and is denoted by $|z|$. We see

that

$$|z| = |a + bi| = \sqrt{a^2 + b^2}. \tag{7.8.8}$$

There is another concept related to complex number that is based on the following bit of algebra.

$$(a + bi)(a - bi) = a^2 - (bi)^2 = a^2 - b^2 i^2 = a^2 + b^2 \tag{7.8.9}$$

The complex number $a - bi$ is called the complex conjugate of $a + bi$. If we let $z = a + bi$, we denote the **complex conjugate** of z as $\overline{z}$. So

$$\overline{z} = \overline{a + bi} = a - bi. \tag{7.8.10}$$

We also notice that

$$z\overline{z} = (a + bi)(a - bi) = a^2 + b^2, \tag{7.8.11}$$

and so the product of a complex number with its conjugate is a real number. In fact,

$$z\overline{z} = a^2 + b^2 = |z|^2, \tag{7.8.12}$$

, and so

$$|z| = \sqrt{z\overline{z}} \tag{7.8.13}$$

Exercise 7.8.5

Let $w = 2 + 3i$ and $z = -1 + 5i$

a. Find $\overline{w}$ and $\overline{z}$.
b. Compute $|w|$ and $|z|$.
c. Compute $w\overline{w}$ and $z\overline{z}$.
d. What is $\overline{z}$ if z is a real number?

Answer available at the end of this section.

Summary

In this section, we studied the following important concepts and ideas:

- A **complex number** is an object of the form $a + bi$, where a and b are real numbers and $i^2 = -1$. When we have a complex number of the form $z = a + bi$, the number a is called the **real part** of the complex number z and the number b is called the **imaginary part** of z.
- We can add, subtract, multiply, and divide complex numbers as follows:

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

$$(a + bi) - (c + di) = (a - c) + (b - d)i$$

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i$$

$$\frac{a + bi}{c + di} = \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}i$$

provided $c + di \neq 0$

- A complex number $a + bi$ can be represented geometrically with a vector in standard position with terminal point (a, b) . When we do this, the horizontal axis is called **the real axis**, and the vertical axis is called **the imaginary axis**. In addition, the coordinate plane is then referred to as **the complex plane**. That is, if $z = a + bi$ we can think of z as a directed line segment from the origin to the point (a, b), where the terminal point of the segment is a units from the imaginary axis and b units from the real axis.
- The **norm** (or **modulus**) of the complex number $z = a + bi$ is the distance from the origin to the point (a, b) and is denoted by $|z|$. We see that

$$|z| = |a + bi| = \sqrt{a^2 + b^2}$$

- The complex number $a - bi$ is called the complex conjugate of $a + bi$. Note that

$$(a + bi)(a - bi) = a^2 + b^2 = |a + bi|^2$$

Answers

Exercise $7.8.1A$

(a) $(2 + 3i) + (7 - 4i) = 9 - i$
(b) $(4 - 2i)(3 + i) = (4 - 2i)3 + (4 - 2i)i = 14 - 2i$
(c) $(2 + i)i - (3 + 4i) = (2i - 1) - 3 - 4i = -4 - 2i$

Exercise $7.8.1B$

We use the quadratic formula to solve the equation and obtain

$$x = \frac{1 \pm \sqrt{-7}}{2}. \tag{7.8.14}$$

We can then write $\sqrt{-7} = i\sqrt{7}$. So the two solutions of the quadratic equation are:

$$\begin{aligned} x &= \frac{1 \pm i\sqrt{7}}{2} \\ &= \frac{1}{2} \pm \frac{\sqrt{7}}{2} i \end{aligned}$$

Exercise 7.8.3

a. Using our formula with $a = 5, b = -1, c = 3$ and $d = 4$ gives us

$$\frac{5-i}{3+4i} = \frac{15-4}{15} + \frac{-3-20}{25}i = \frac{11}{25} - \frac{23}{25}i \tag{7.8.15}$$

As a check, we see that

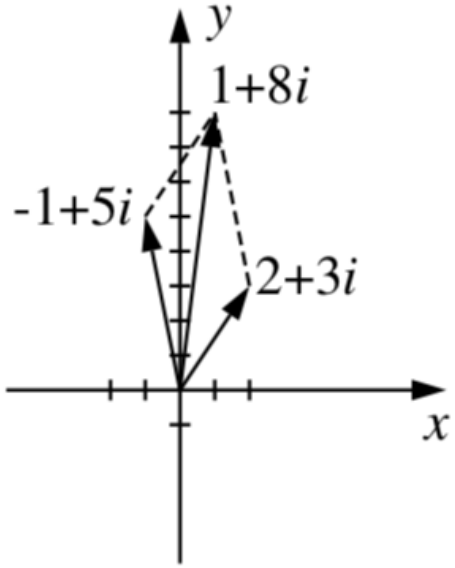
$$\left(\frac{11}{25} - \frac{23}{25}i\right)(3+4i) = \left(\frac{33}{25} - \frac{69}{25}i\right) + \frac{44}{25}i - \frac{92}{25}i^2 = \left(\frac{33}{25} + \frac{92}{25}\right) + \left(-\frac{69}{25}i + \frac{44}{25}i\right) = 5 - i \tag{7.8.16}$$

b. We can solve for x by dividing both sides of the equation by $3 + 4i$ to see that

$$x = \frac{5-i}{3+4i} = \frac{11}{25} - \frac{23}{25}i \tag{7.8.17}$$

Exercise 7.8.4

1. The sum is $w + z = (2 - 1) + (3 + 5)i = 1 + 8i$.
2. A representation of the complex sum using vectors is shown in the figure below.



Exercise 7.8.5

1. Using the definition of the conjugate of a complex number we find that $\overline{w} = 2 - 3i$ and $\overline{z} = -1 - 5i$.
2. Using the definition of the norm of a complex number we find that $|w| = \sqrt{2^2 + 3^2} = \sqrt{13}$ and $|z| = \sqrt{(-1)^2 + 5^2} = \sqrt{26}$.
3. Using the definition of the product of complex numbers we find that

$$w\overline{w} = (2 + 3i)(2 - 3i) = 4 + 9 = 13 \tag{7.8.18}$$

$$z\overline{z} = (-1 + 5i)(-1 - 5i) = 1 + 25 = 26 \tag{7.8.19}$$

4. Let $z = a + 0i$ for some $a \,^{\text{int}}\mathbb{R}$. Then $\overline{z} = a - 0i$. Thus, $\overline{z} = z$ when $z \,^{\text{int}}\mathbb{R}$.

This page titled 7.8: The Complex Number System is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Ted Sundstrom and Steven Schlicker (ScholarWorks @Grand Valley State University) via source content that was edited to the style and standards of the LibreTexts platform.

- 5.1: The Complex Number System by Ted Sundstrom and Steven Schlicker is licensed CC BY-NC-SA 3.0. Original source: https://scholarworks.gvsu.edu/books/12.

7.8.1: Solve Systems of Linear Equations with Complex Coefficients

In many applications, especially in engineering, physics, and applied mathematics, systems of linear equations may involve **complex numbers** as coefficients or constants. Solving such systems follows the same principles as real-valued systems, but requires careful handling of the real and imaginary parts.

Read this paper

https://www.math.uwaterloo.ca/~jmcki.../Lecture3i.pdf

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This page titled 7.8.1: Solve Systems of Linear Equations with Complex Coefficients is shared under a license and was authored, remixed, and/or curated by LibreTexts.

INDEX

A

absorbing Markov chain
7.2.1: Absorbing Markov Chains

adjugate
3.3E: Exercises for Section 3.3

adjugate matrix
3.3: Application of the Determinant to Inverses; Cramer's Rule

algebra
4.4E: Exercises for Section 4.3
4.5: Spanning Sets in R^n

algorithm
4.12: Gram-Schmidt Process
7.1: Stochastic Matrices

angle
4.10: Dot Products and Orthogonality

angles
4.3E: Exercises for Section 4.2
5.1: Prelude to Linear Transformations and Matrix Algebra

approximating_points
4.16.E: Exercises for Section 4.11

approximation
7.3: The Method of Least Squares

area
3.4: Determinants and Geometry
3.4E: Exercises for Section 3.4
4.3E: Exercises for Section 4.2

B

basis
4.11: Orthogonal Vectors and Matrices
4.12.E: Exercises for Section 4.9
4.14: Orthogonal Projections
4.14.E: Exercise for Section 4.10
4.15: Orthogonal Projection
4.5: Spanning Sets in R^n
4.8: Subspaces and Bases
4.8.E: Exercise for Section 4.6
4.9: Row, Column and Null Spaces
4.9.E: Exercise for Section 4.7

best_fit
4.16.E: Exercises for Section 4.11

Binet formula
7.4: An Application to Linear Recurrences

C

canonical form
7.2.1: Absorbing Markov Chains

CauchySchwarz
4.3: Dot and Cross Product

closest_point
4.14: Orthogonal Projections

codomain
5.3: Matrix Transformations

coefficients
4.6: Linear Independence

cofactor
3.1: Basic Techniques
3.2: Properties of Determinants
3.3E: Exercises for Section 3.3

cofactor matrix
3.3: Application of the Determinant to Inverses; Cramer's Rule

collinear
3.4E: Exercises for Section 3.4

column rank
4.13: Orthogonal Complements

column space
4.9: Row, Column and Null Spaces
4.9.E: Exercise for Section 4.7
5.5: The Invertible Matrix Theorem

complex conjugate
7.8: The Complex Number System

complex numbers
7.8: The Complex Number System

complex plane
7.8: The Complex Number System

components
4.2: Review of Vectors

convergence
7.1: Stochastic Matrices

coordinates
5.1: Prelude to Linear Transformations and Matrix Algebra

Cramer's rule
3: Determinants
3.3E: Exercises for Section 3.3

Cramer's Rule
3.3: Application of the Determinant to Inverses; Cramer's Rule

cross product
4: Vector Spaces - R^n

crossproduct
4.3: Dot and Cross Product
4.3E: Exercises for Section 4.2

D

data analysis
7.3: The Method of Least Squares

data_analysis
4.16: Least Squares Approximation

determinant
3: Determinants
3.1: Basic Techniques
3.2: Properties of Determinants
3.2E: Exercises for Section 3.2
3.4: Determinants and Geometry
3.4E: Exercises for Section 3.4

determinants
3.2: Properties of Determinants
3.3: Application of the Determinant to Inverses; Cramer's Rule
3.3E: Exercises for Section 3.3
4.6.E: Exercise for Section 4.5

diagonal
4.11.E: Exercise for Section 4.8

dilation
5.2: Linear Transformations

dimension
4.8: Subspaces and Bases
4.8.E: Exercise for Section 4.6

dimensionality
4.9.E: Exercise for Section 4.7

dimensions
5.4: One-to-one and Onto Transformations

direction
4.2E: Exercises for Section 4.1

direction vector
4.4: Lines and Planes
4.4E: Exercises for Section 4.3

distance
4.14.E: Exercise for Section 4.10
4.15: Orthogonal Projection
4.2: Review of Vectors

distance minimization
7.3: The Method of Least Squares

distance_to_plane
4.16.E: Exercises for Section 4.11

domain
5.3: Matrix Transformations

dot product
4: Vector Spaces - R^n
4.10: Dot Products and Orthogonality
4.11: Orthogonal Vectors and Matrices

dotproduct
4.3: Dot and Cross Product

drunkard's random walk
7.2.1: Absorbing Markov Chains

E

eigenvalues
7.1: Stochastic Matrices

elementary matrix
2.5: Elementary Matrices

equations
4.5: Spanning Sets in R^n

equilibrium vector
7.2.2: Introduction to Markov Chains

example
4.4: Lines and Planes
4.8: Subspaces and Bases
5.4: One-to-one and Onto Transformations

examples
3.2: Properties of Determinants

exercise
4.11.E: Exercise for Section 4.8
4.12.E: Exercises for Section 4.9
4.5.E: Exercise for Section 4.4

exercises
3: Determinants
4.8.E: Exercise for Section 4.6

experimental_data
4.16.E: Exercises for Section 4.11

F

fixed probability vector
7.2.2: Introduction to Markov Chains

Fourier
4.11.E: Exercise for Section 4.8

Fourier expansion
4.11: Orthogonal Vectors and Matrices

function
5.4: One-to-one and Onto Transformations

G

gamber's ruin problem
7.2.1: Absorbing Markov Chains
geometric interpretation
4.5.E: Exercise for Section 4.4
geometric representation
7.3: The Method of Least Squares
geometric_shapes
4.2E: Exercises for Section 4.1
geometry
3: Determinants
3.4: Determinants and Geometry
4.14.E: Exercise for Section 4.10
4.16.E: Exercises for Section 4.11
4.3: Dot and Cross Product
4.3E: Exercises for Section 4.2
4.4: Lines and Planes
4.4E: Exercises for Section 4.3
4.5.E: Exercise for Section 4.4
5.1: Prelude to Linear Transformations and Matrix Algebra
5.2: Linear Transformations
5.3: Matrix Transformations
Google matrix
7.1: Stochastic Matrices

I

identity matrix
2.3: The Identity and Inverses
image
4.9.E: Exercise for Section 4.7
independence
4.5: Spanning Sets in R^n
injective
5.4: One-to-one and Onto Transformations
inner product
4: Vector Spaces - R^n
inverse
3.3E: Exercises for Section 3.3
inverse calculation
3.3: Application of the Determinant to Inverses; Cramer's Rule
inverse kinematics
5.1: Prelude to Linear Transformations and Matrix Algebra
inverses
3: Determinants
invertible
3.3E: Exercises for Section 3.3
5.5: The Invertible Matrix Theorem
invertible matrix theorem
5: Linear Transformations

J

joints
5.1: Prelude to Linear Transformations and Matrix Algebra

K

kernel
4.9.E: Exercise for Section 4.7
Kronecker delta
2.3: The Identity and Inverses

L

Laplace Expansion
3.1: Basic Techniques
Law of Cosines
4.10: Dot Products and Orthogonality
learning
4.2: Review of Vectors
4.5: Spanning Sets in R^n
4.9: Row, Column and Null Spaces
learning objectives
4.13: Orthogonal Complements
4.15: Orthogonal Projection
least squares
4.16: Least Squares Approximation
least_squares
4.16.E: Exercises for Section 4.11
line
4.4: Lines and Planes
4.4E: Exercises for Section 4.3
linear
4.5: Spanning Sets in R^n
linear algebra
3.2: Properties of Determinants
3.2E: Exercises for Section 3.2
3.3: Application of the Determinant to Inverses; Cramer's Rule
3.3E: Exercises for Section 3.3
3.4: Determinants and Geometry
4.11: Orthogonal Vectors and Matrices
4.13: Orthogonal Complements
4.15: Orthogonal Projection
4.2: Review of Vectors
4.4: Lines and Planes
4.5.E: Exercise for Section 4.4
4.6: Linear Independence
4.9: Row, Column and Null Spaces
5: Linear Transformations
5.3: Matrix Transformations
5.4: One-to-one and Onto Transformations
5.5: The Invertible Matrix Theorem
linear combinations
4: Vector Spaces - R^n
4.12: Gram-Schmidt Process
4.5.E: Exercise for Section 4.4
4.8: Subspaces and Bases
linear dependence
4.6: Linear Independence
linear equations
4.16: Least Squares Approximation
4.4E: Exercises for Section 4.3
7.3: The Method of Least Squares
linear independence
4: Vector Spaces - R^n
4.11: Orthogonal Vectors and Matrices
4.12: Gram-Schmidt Process
4.6: Linear Independence
4.8: Subspaces and Bases
4.9.E: Exercise for Section 4.7
5.5: The Invertible Matrix Theorem
Linear Systems
3.3: Application of the Determinant to Inverses; Cramer's Rule
linear transformation
5.2: Linear Transformations
linear transformations
4.15: Orthogonal Projection
4.9.E: Exercise for Section 4.7
5: Linear Transformations
linear_algebra
4.14: Orthogonal Projections
4.14.E: Exercise for Section 4.10
4.2E: Exercises for Section 4.1
4.8.E: Exercise for Section 4.6
linear_combination
4.2E: Exercises for Section 4.1
linear_equation
4.16.E: Exercises for Section 4.11
linear_independence
4.8.E: Exercise for Section 4.6
LinearAlgebra
4.3: Dot and Cross Product
4.6.E: Exercise for Section 4.5
linearcombination
4.6.E: Exercise for Section 4.5
linearindependence
4.6.E: Exercise for Section 4.5
linearly dependent
3.4E: Exercises for Section 3.4
lines
4: Vector Spaces - R^n
LU factorization
2.6: LU Factorization

M

magnitude
4.2: Review of Vectors
4.2E: Exercises for Section 4.1
main diagonal
2.1: Matrix Addition, Scalar Multiplication, and Transposition
Markov chains
7.1: Stochastic Matrices
7.2: An Application to Markov Chains
7.2.2: Introduction to Markov Chains
Markov process
7.2.2: Introduction to Markov Chains
Math
4.3: Dot and Cross Product
mathematical concepts
4.13: Orthogonal Complements
mathematics
4.5: Spanning Sets in R^n
4.5.E: Exercise for Section 4.4
matrices
3.3E: Exercises for Section 3.3
4.8.E: Exercise for Section 4.6
5: Linear Transformations
5.1: Prelude to Linear Transformations and Matrix Algebra
matrix
2.1: Matrix Addition, Scalar Multiplication, and Transposition
3.2: Properties of Determinants
3.2E: Exercises for Section 3.2
3.4: Determinants and Geometry
3.4E: Exercises for Section 3.4
4.11: Orthogonal Vectors and Matrices
4.11.E: Exercise for Section 4.8
4.15: Orthogonal Projection
4.16: Least Squares Approximation
4.6: Linear Independence
4.6.E: Exercise for Section 4.5
4.9: Row, Column and Null Spaces
5.2: Linear Transformations
5.3: Matrix Transformations
5.4: One-to-one and Onto Transformations
5.5: The Invertible Matrix Theorem
7.3: The Method of Least Squares

matrix addition
2.1: Matrix Addition, Scalar Multiplication, and Transposition
matrix algebra
5: Linear Transformations
matrix form
4.10: Dot Products and Orthogonality
matrix inverse
2.4: Finding the Inverse of a Matrix
3.3: Application of the Determinant to Inverses; Cramer's Rule
matrix inverses
5: Linear Transformations
matrix multiplication
4.13: Orthogonal Complements
5: Linear Transformations
matrix rank
4.9.E: Exercise for Section 4.7
methods_of_solution
3.3: Application of the Determinant to Inverses; Cramer's Rule
minor of a matrix
3.1: Basic Techniques
minors or a matrix
3.1: Basic Techniques
motion
5.1: Prelude to Linear Transformations and Matrix Algebra
multivariable calculus
3: Determinants

N

nilpotent
3.2E: Exercises for Section 3.2
nonnegative matrix
7.1: Stochastic Matrices
normal equation
4.16: Least Squares Approximation
null space
4.13: Orthogonal Complements
4.9: Row, Column and Null Spaces
nullity
4.9: Row, Column and Null Spaces
5.5: The Invertible Matrix Theorem

O

objectives
4.2: Review of Vectors
4.5: Spanning Sets in R^n
4.9: Row, Column and Null Spaces
onto
5.4: One-to-one and Onto Transformations
onto transformations
5: Linear Transformations
optimization
4.16: Least Squares Approximation
7.3: The Method of Least Squares
orthogonal
3.2E: Exercises for Section 3.2
4.10: Dot Products and Orthogonality
4.11.E: Exercise for Section 4.8
4.12: Gram-Schmidt Process
4.12.E: Exercises for Section 4.9
orthogonal decomposition
4.15: Orthogonal Projection
orthogonal projection
4.15: Orthogonal Projection
4.16: Least Squares Approximation
orthogonal vectors
4.11: Orthogonal Vectors and Matrices
orthogonal_complement
4.14: Orthogonal Projections
4.14.E: Exercise for Section 4.10
orthogonal_projection
4.14: Orthogonal Projections
4.14.E: Exercise for Section 4.10
orthogonality
4.10: Dot Products and Orthogonality
orthonormal
4.11: Orthogonal Vectors and Matrices
4.11.E: Exercise for Section 4.8
4.12: Gram-Schmidt Process
4.12.E: Exercises for Section 4.9
orthonormal_basis
4.14: Orthogonal Projections

P

PageRank
7.1: Stochastic Matrices
parallelepiped
3.4: Determinants and Geometry
4.3E: Exercises for Section 4.2
parallelogram
3.4: Determinants and Geometry
3.4E: Exercises for Section 3.4
4.2E: Exercises for Section 4.1
4.3E: Exercises for Section 4.2
parametric equations
4.4: Lines and Planes
4.4E: Exercises for Section 4.3
parametric form
4.10: Dot Products and Orthogonality
perpendicular
4.10: Dot Products and Orthogonality
plane
4.14.E: Exercise for Section 4.10
4.4: Lines and Planes
planes
4: Vector Spaces - R^n
point
4.4E: Exercises for Section 4.3
polynomial_fit
4.16.E: Exercises for Section 4.11
population dynamics
7.1: Stochastic Matrices
positioning
5.1: Prelude to Linear Transformations and Matrix Algebra
probabilities
7.2: An Application to Markov Chains
products
4.11.E: Exercise for Section 4.8
proof
4.12: Gram-Schmidt Process
4.14.E: Exercise for Section 4.10
4.6.E: Exercise for Section 4.5
4.8: Subspaces and Bases
5.2: Linear Transformations
proofs
3.2: Properties of Determinants
3.2E: Exercises for Section 3.2
properties
3: Determinants
3.2: Properties of Determinants
5.2: Linear Transformations
properties of determinants
3.2E: Exercises for Section 3.2

R

R3
4.4: Lines and Planes
range
5.3: Matrix Transformations
rank
4.9: Row, Column and Null Spaces
5.5: The Invertible Matrix Theorem
real_numbers
4.8.E: Exercise for Section 4.6
reflection
5.3: Matrix Transformations
regression
4.16: Least Squares Approximation
7.3: The Method of Least Squares
regular Markov chains
7.2.3: Regular Markov Chains
regular stochastic matrix
7.1: Stochastic Matrices
residual
4.16: Least Squares Approximation
Rn
4.2: Review of Vectors
4.4: Lines and Planes
robot arm
5.1: Prelude to Linear Transformations and Matrix Algebra
rotation
5.2: Linear Transformations
5.3: Matrix Transformations
row echelon form
5.5: The Invertible Matrix Theorem
row operations
3: Determinants
3.2: Properties of Determinants
3.2E: Exercises for Section 3.2
3.4: Determinants and Geometry
row rank
4.13: Orthogonal Complements
row space
4.9: Row, Column and Null Spaces
4.9.E: Exercise for Section 4.7

S

scalar multiplication
4.2: Review of Vectors
scalar product
4: Vector Spaces - R^n
scalarproduct
4.3: Dot and Cross Product
scalars
4.6.E: Exercise for Section 4.5
shear
5.3: Matrix Transformations
similarity
3.2E: Exercises for Section 3.2
skew symmetric
3.4E: Exercises for Section 3.4

skewsymmetricalmatrix
4.3E: Exercises for Section 4.2
solution
4.16.E: Exercises for Section 4.11
solution matrix
7.2.1: Absorbing Markov Chains
solutions
4.6: Linear Independence
span
4.10: Dot Products and Orthogonality
4.11.E: Exercise for Section 4.8
4.12: Gram-Schmidt Process
4.12.E: Exercises for Section 4.9
4.13: Orthogonal Complements
4.14: Orthogonal Projections
4.5.E: Exercise for Section 4.4
4.6: Linear Independence
4.6.E: Exercise for Section 4.5
4.8: Subspaces and Bases
square matrix
3: Determinants
state vector
7.2.2: Introduction to Markov Chains
statistics
4.16: Least Squares Approximation
stochastic matrices
7.1: Stochastic Matrices
subspace
4.11: Orthogonal Vectors and Matrices
4.12: Gram-Schmidt Process
4.12.E: Exercises for Section 4.9
4.13: Orthogonal Complements
4.14: Orthogonal Projections
4.15: Orthogonal Projection
4.8: Subspaces and Bases
subspaces
4: Vector Spaces - R^n
4.14.E: Exercise for Section 4.10
4.5: Spanning Sets in R^n
4.8.E: Exercise for Section 4.6
symmetric
4.11.E: Exercise for Section 4.8
system_of_equations
4.16.E: Exercises for Section 4.11
systems of equations
3.3E: Exercises for Section 3.3
4.6: Linear Independence
systems_of_equations
3.3: Application of the Determinant to Inverses; Cramer's Rule

T

theorem
4.8: Subspaces and Bases
5.5: The Invertible Matrix Theorem
theorems
4.6: Linear Independence
transformation
4.11: Orthogonal Vectors and Matrices
4.12.E: Exercises for Section 4.9
5.1: Prelude to Linear Transformations and Matrix Algebra
5.3: Matrix Transformations
transformation properties
5.2: Linear Transformations
transformations
5.4: One-to-one and Onto Transformations
transition matrix
7.2.2: Introduction to Markov Chains
transpose
3.2: Properties of Determinants
3.2E: Exercises for Section 3.2
triangle
3.4E: Exercises for Section 3.4
4.4E: Exercises for Section 4.3
triangleinequality
4.3: Dot and Cross Product
triangular matrix
3.1: Basic Techniques

U

unique solution
5.4: One-to-one and Onto Transformations
5.5: The Invertible Matrix Theorem
upper triangular
3.3E: Exercises for Section 3.3

V

vector
4.13: Orthogonal Complements
4.14: Orthogonal Projections
4.14.E: Exercise for Section 4.10
4.15: Orthogonal Projection
4.2: Review of Vectors
4.4: Lines and Planes
5.2: Linear Transformations
vector addition
4.2: Review of Vectors
vector equation
4.4E: Exercises for Section 4.3
vector space
4.5.E: Exercise for Section 4.4
vector spaces
4.9.E: Exercise for Section 4.7
5.2: Linear Transformations
vector_operations
4.2E: Exercises for Section 4.1
vector_space
4.6.E: Exercise for Section 4.5
vector_spaces
4.8.E: Exercise for Section 4.6
vectorprojections
4.3: Dot and Cross Product
vectors
3.4: Determinants and Geometry
3.4E: Exercises for Section 3.4
4: Vector Spaces - R^n
4.10: Dot Products and Orthogonality
4.11.E: Exercise for Section 4.8
4.12: Gram-Schmidt Process
4.2E: Exercises for Section 4.1
4.3: Dot and Cross Product
4.3E: Exercises for Section 4.2
4.5: Spanning Sets in R^n
4.5.E: Exercise for Section 4.4
4.6: Linear Independence
4.6.E: Exercise for Section 4.5
4.8: Subspaces and Bases
4.8.E: Exercise for Section 4.6
volume
3.4: Determinants and Geometry
3.4E: Exercises for Section 3.4
4.3E: Exercises for Section 4.2
volumes
3: Determinants

Z

zero vector
4.5.E: Exercise for Section 4.4

This page titled Index is shared under a license and was authored, remixed, and/or curated by LibreTexts.

GLOSSARY

S

Sample Word 1 | Sample Definition 1

This page titled Glossary is shared under a license and was authored, remixed, and/or curated by LibreTexts.

DETAILED LICENSING

Overview

Title: Linear Algebra and Its Application

Webpages: 101

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY 4.0: 50.5% (51 pages)
- Undeclared: 29.7% (30 pages)
- GNU Free Documentation License: 12.9% (13 pages)
- CC BY-NC-SA 4.0: 5.9% (6 pages)
- CC BY-NC-SA 3.0: 1% (1 page)

By Page

- Linear Algebra and Its Application - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 0: Why Linear Algebra? - *Undeclared*
  - 1: Solving Systems of Linear Equations - *CC BY 4.0*
    - 1.1: Systems of Linear Equations - *Undeclared*
    - 1.2: Matrix Notation and Row Reduction - *Undeclared*
    - 1.3: Rank and Homogeneous Systems - *Undeclared*
    - 1.4: Chemistry Reaction-Balancing Chemical Reaction - *Undeclared*
    - 1.5: Applications to Physics and Resistor Networks - *Undeclared*
    - 1.6: Chapters Exercises - *Undeclared*
  - 2: Matrices - *CC BY 4.0*
    - 2.1: Matrix Addition, Scalar Multiplication, and Transposition - *CC BY-NC-SA 4.0*
    - 2.2: Matrix Multiplication - *Undeclared*
    - 2.3: The Identity and Inverses - *CC BY 4.0*
    - 2.4: Finding the Inverse of a Matrix - *CC BY 4.0*
    - 2.5: Elementary Matrices - *CC BY 4.0*
    - 2.6: LU Factorization - *CC BY 4.0*
  - 3: Determinants - *CC BY 4.0*
    - 3.1: Basic Techniques - *CC BY 4.0*
      - 3.1E: Exercises for Section 3.1 - *CC BY 4.0*
    - 3.2: Properties of Determinants - *CC BY 4.0*
      - 3.2E: Exercises for Section 3.2 - *CC BY 4.0*
    - 3.3: Application of the Determinant to Inverses; Cramer's Rule - *CC BY 4.0*
      - 3.3E: Exercises for Section 3.3 - *CC BY 4.0*
    - 3.4: Determinants and Geometry - *CC BY 4.0*
      - 3.4E: Exercises for Section 3.4 - *CC BY 4.0*
  - 4: Vector Spaces - R^n - *CC BY 4.0*
    - 4.1: Subspaces - *GNU Free Documentation License*
    - 4.2: Review of Vectors - *CC BY 4.0*
      - 4.2E: Exercises for Section 4.1 - *CC BY 4.0*
    - 4.3: Dot and Cross Product - *CC BY 4.0*
      - 4.3E: Exercises for Section 4.2 - *CC BY 4.0*
    - 4.4: Lines and Planes - *CC BY 4.0*
      - 4.4E: Exercises for Section 4.3 - *CC BY 4.0*
    - 4.5: Spanning Sets in R^n - *CC BY 4.0*
      - 4.5.E: Exercise for Section 4.4 - *CC BY 4.0*
    - 4.6: Linear Independence - *CC BY 4.0*
      - 4.6.E: Exercise for Section 4.5 - *CC BY 4.0*
    - 4.7: Subspaces - *GNU Free Documentation License*
    - 4.8: Subspaces and Bases - *CC BY 4.0*
      - 4.8.E: Exercise for Section 4.6 - *CC BY 4.0*
    - 4.9: Row, Column and Null Spaces - *CC BY 4.0*
      - 4.9.E: Exercise for Section 4.7 - *CC BY 4.0*
    - 4.10: Dot Products and Orthogonality - *GNU Free Documentation License*
    - 4.11: Orthogonal Vectors and Matrices - *CC BY 4.0*
      - 4.11.E: Exercise for Section 4.8 - *CC BY 4.0*
    - 4.12: Gram-Schmidt Process - *CC BY 4.0*
      - 4.12.E: Exercises for Section 4.9 - *CC BY 4.0*
    - 4.13: Orthogonal Complements - *GNU Free Documentation License*
    - 4.14: Orthogonal Projections - *CC BY 4.0*
      - 4.14.E: Exercise for Section 4.10 - *CC BY 4.0*
    - 4.15: Orthogonal Projection - *GNU Free Documentation License*

- 4.16: Least Squares Approximation - *CC BY 4.0*
  - 4.16.E: Exercises for Section 4.11 - *CC BY 4.0*
- 4.17: Vector Spaces and Subspace - *Undeclared*
- 4.18: Amira Vector Space Axioms - *Undeclared*
- 4.19: Amira Vector Subspace of R^n - *Undeclared*
- 4.20: The Real Vector Space ^R^n - *Undeclared*
- 4.E: Exercises - *CC BY 4.0*

◦ 5: Linear Transformations - *GNU Free Documentation License*
- 5.1: Prelude to Linear Transformations and Matrix Algebra - *GNU Free Documentation License*
- 5.2: Linear Transformations - *GNU Free Documentation License*
- 5.3: Matrix Transformations - *GNU Free Documentation License*
- 5.4: One-to-one and Onto Transformations - *GNU Free Documentation License*
- 5.5: The Invertible Matrix Theorem - *GNU Free Documentation License*

◦ 6: Eigenvalues and Eigenvectors - *CC BY 4.0*
- 6.1: An introduction to Eigenvalues and Eigenvectors - *CC BY 4.0*
- 6.2: Finding eigenvalues and eigenvectors - *CC BY 4.0*
- 6.3: Diagonalization, similarity, and powers of a matrix - *CC BY 4.0*
- 6.4: Dynamical Systems - *CC BY 4.0*
- 6.5: Markov chains and Google's PageRank algorithm - *CC BY 4.0*

◦ 7: Honors Projects - *Undeclared*
- 7.1: Stochastic Matrices - *GNU Free Documentation License*
- 7.2: An Application to Markov Chains - *CC BY-NC-SA 4.0*
  - 7.2.1: Absorbing Markov Chains - *CC BY 4.0*
    - 7.2.1.1: Absorbing Markov Chains (Exercises) - *CC BY 4.0*
  - 7.2.2: Introduction to Markov Chains - *CC BY 4.0*
    - 7.2.2.1: Introduction to Markov Chains (Exercises) - *CC BY 4.0*
  - 7.2.3: Regular Markov Chains - *CC BY 4.0*
    - 7.2.3.1: Regular Markov Chains (Exercises) - *CC BY 4.0*
- 7.3: The Method of Least Squares - *GNU Free Documentation License*
- 7.4: An Application to Linear Recurrences - *CC BY-NC-SA 4.0*
  - 7.4E: An Application to Linear Recurrences Exercises - *Undeclared*
- 7.5: An Application to Systems of Differential Equations - *CC BY-NC-SA 4.0*
  - 7.5E: An Application to Systems of Differential Equations Exercises - *Undeclared*
- 7.6: An Application to Computer Graphics - *CC BY-NC-SA 4.0*
  - 7.6E: An Application to Computer Graphics Exercises - *Undeclared*
- 7.7: An Application to Input-Output Economic Models - *CC BY-NC-SA 4.0*
  - 7.7E: An Application to Input-Output Economic Models Exercises - *Undeclared*
- 7.8: The Complex Number System - *CC BY-NC-SA 3.0*
  - 7.8.1: Solve Systems of Linear Equations with Complex Coefficients - *Undeclared*

◦ Back Matter - *Undeclared*
- Index - *Undeclared*
- Glossary - *Undeclared*
- Detailed Licensing - *Undeclared*
- Detailed Licensing - *Undeclared*
- Detailed Licensing - *Undeclared*

This page titled Linear Algebra and Its Application is shared under a license and was authored, remixed, and/or curated by Amira Alkeswani, Cañada College.

DETAILED LICENSING

Overview

Title: Linear Algebra and Its Application

Webpages: 101

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY 4.0: 50.5% (51 pages)
- Undeclared: 29.7% (30 pages)
- GNU Free Documentation License: 12.9% (13 pages)
- CC BY-NC-SA 4.0: 5.9% (6 pages)
- CC BY-NC-SA 3.0: 1% (1 page)

By Page

- Linear Algebra and Its Application - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 0: Why Linear Algebra? - *Undeclared*
  - 1: Solving Systems of Linear Equations - *CC BY 4.0*
    - 1.1: Systems of Linear Equations - *Undeclared*
    - 1.2: Matrix Notation and Row Reduction - *Undeclared*
    - 1.3: Rank and Homogeneous Systems - *Undeclared*
    - 1.4: Chemistry Reaction-Balancing Chemical Reaction - *Undeclared*
    - 1.5: Applications to Physics and Resistor Networks - *Undeclared*
    - 1.6: Chapters Exercises - *Undeclared*
  - 2: Matrices - *CC BY 4.0*
    - 2.1: Matrix Addition, Scalar Multiplication, and Transposition - *CC BY-NC-SA 4.0*
    - 2.2: Matrix Multiplication - *Undeclared*
    - 2.3: The Identity and Inverses - *CC BY 4.0*
    - 2.4: Finding the Inverse of a Matrix - *CC BY 4.0*
    - 2.5: Elementary Matrices - *CC BY 4.0*
    - 2.6: LU Factorization - *CC BY 4.0*
  - 3: Determinants - *CC BY 4.0*
    - 3.1: Basic Techniques - *CC BY 4.0*
      - 3.1E: Exercises for Section 3.1 - *CC BY 4.0*
    - 3.2: Properties of Determinants - *CC BY 4.0*
      - 3.2E: Exercises for Section 3.2 - *CC BY 4.0*
    - 3.3: Application of the Determinant to Inverses; Cramer's Rule - *CC BY 4.0*
      - 3.3E: Exercises for Section 3.3 - *CC BY 4.0*
    - 3.4: Determinants and Geometry - *CC BY 4.0*
      - 3.4E: Exercises for Section 3.4 - *CC BY 4.0*
  - 4: Vector Spaces - R^n - *CC BY 4.0*
    - 4.1: Subspaces - *GNU Free Documentation License*
    - 4.2: Review of Vectors - *CC BY 4.0*
      - 4.2E: Exercises for Section 4.1 - *CC BY 4.0*
    - 4.3: Dot and Cross Product - *CC BY 4.0*
      - 4.3E: Exercises for Section 4.2 - *CC BY 4.0*
    - 4.4: Lines and Planes - *CC BY 4.0*
      - 4.4E: Exercises for Section 4.3 - *CC BY 4.0*
    - 4.5: Spanning Sets in R^n - *CC BY 4.0*
      - 4.5.E: Exercise for Section 4.4 - *CC BY 4.0*
    - 4.6: Linear Independence - *CC BY 4.0*
      - 4.6.E: Exercise for Section 4.5 - *CC BY 4.0*
    - 4.7: Subspaces - *GNU Free Documentation License*
    - 4.8: Subspaces and Bases - *CC BY 4.0*
      - 4.8.E: Exercise for Section 4.6 - *CC BY 4.0*
    - 4.9: Row, Column and Null Spaces - *CC BY 4.0*
      - 4.9.E: Exercise for Section 4.7 - *CC BY 4.0*
    - 4.10: Dot Products and Orthogonality - *GNU Free Documentation License*
    - 4.11: Orthogonal Vectors and Matrices - *CC BY 4.0*
      - 4.11.E: Exercise for Section 4.8 - *CC BY 4.0*
    - 4.12: Gram-Schmidt Process - *CC BY 4.0*
      - 4.12.E: Exercises for Section 4.9 - *CC BY 4.0*
    - 4.13: Orthogonal Complements - *GNU Free Documentation License*
    - 4.14: Orthogonal Projections - *CC BY 4.0*
      - 4.14.E: Exercise for Section 4.10 - *CC BY 4.0*
    - 4.15: Orthogonal Projection - *GNU Free Documentation License*

- 4.16: Least Squares Approximation - *CC BY 4.0*
  - 4.16.E: Exercises for Section 4.11 - *CC BY 4.0*
- 4.17: Vector Spaces and Subspace - *Undeclared*
- 4.18: Amira Vector Space Axioms - *Undeclared*
- 4.19: Amira Vector Subspace of R^n - *Undeclared*
- 4.20: The Real Vector Space ^R^n - *Undeclared*
- 4.E: Exercises - *CC BY 4.0*

◦ 5: Linear Transformations - *GNU Free Documentation License*
- 5.1: Prelude to Linear Transformations and Matrix Algebra - *GNU Free Documentation License*
- 5.2: Linear Transformations - *GNU Free Documentation License*
- 5.3: Matrix Transformations - *GNU Free Documentation License*
- 5.4: One-to-one and Onto Transformations - *GNU Free Documentation License*
- 5.5: The Invertible Matrix Theorem - *GNU Free Documentation License*

◦ 6: Eigenvalues and Eigenvectors - *CC BY 4.0*
- 6.1: An introduction to Eigenvalues and Eigenvectors - *CC BY 4.0*
- 6.2: Finding eigenvalues and eigenvectors - *CC BY 4.0*
- 6.3: Diagonalization, similarity, and powers of a matrix - *CC BY 4.0*
- 6.4: Dynamical Systems - *CC BY 4.0*
- 6.5: Markov chains and Google's PageRank algorithm - *CC BY 4.0*

◦ 7: Honors Projects - *Undeclared*
- 7.1: Stochastic Matrices - *GNU Free Documentation License*
- 7.2: An Application to Markov Chains - *CC BY-NC-SA 4.0*
  - 7.2.1: Absorbing Markov Chains - *CC BY 4.0*
    - 7.2.1.1: Absorbing Markov Chains (Exercises) - *CC BY 4.0*
  - 7.2.2: Introduction to Markov Chains - *CC BY 4.0*
    - 7.2.2.1: Introduction to Markov Chains (Exercises) - *CC BY 4.0*
  - 7.2.3: Regular Markov Chains - *CC BY 4.0*
    - 7.2.3.1: Regular Markov Chains (Exercises) - *CC BY 4.0*
- 7.3: The Method of Least Squares - *GNU Free Documentation License*
- 7.4: An Application to Linear Recurrences - *CC BY-NC-SA 4.0*
  - 7.4E: An Application to Linear Recurrences Exercises - *Undeclared*
- 7.5: An Application to Systems of Differential Equations - *CC BY-NC-SA 4.0*
  - 7.5E: An Application to Systems of Differential Equations Exercises - *Undeclared*
- 7.6: An Application to Computer Graphics - *CC BY-NC-SA 4.0*
  - 7.6E: An Application to Computer Graphics Exercises - *Undeclared*
- 7.7: An Application to Input-Output Economic Models - *CC BY-NC-SA 4.0*
  - 7.7E: An Application to Input-Output Economic Models Exercises - *Undeclared*
- 7.8: The Complex Number System - *CC BY-NC-SA 3.0*
  - 7.8.1: Solve Systems of Linear Equations with Complex Coefficients - *Undeclared*

◦ Back Matter - *Undeclared*
- Index - *Undeclared*
- Glossary - *Undeclared*
- Detailed Licensing - *Undeclared*
- Detailed Licensing - *Undeclared*
- Detailed Licensing - *Undeclared*

This page titled Linear Algebra and Its Application is shared under a license and was authored, remixed, and/or curated by Amira Alkeswani, Cañada College.

DETAILED LICENSING

Overview

Title: Linear Algebra and Its Application

Webpages: 101

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY 4.0: 50.5% (51 pages)
- Undeclared: 29.7% (30 pages)
- GNU Free Documentation License: 12.9% (13 pages)
- CC BY-NC-SA 4.0: 5.9% (6 pages)
- CC BY-NC-SA 3.0: 1% (1 page)

By Page

- Linear Algebra and Its Application - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 0: Why Linear Algebra? - *Undeclared*
  - 1: Solving Systems of Linear Equations - *CC BY 4.0*
    - 1.1: Systems of Linear Equations - *Undeclared*
    - 1.2: Matrix Notation and Row Reduction - *Undeclared*
    - 1.3: Rank and Homogeneous Systems - *Undeclared*
    - 1.4: Chemistry Reaction-Balancing Chemical Reaction - *Undeclared*
    - 1.5: Applications to Physics and Resistor Networks - *Undeclared*
    - 1.6: Chapters Exercises - *Undeclared*
  - 2: Matrices - *CC BY 4.0*
    - 2.1: Matrix Addition, Scalar Multiplication, and Transposition - *CC BY-NC-SA 4.0*
    - 2.2: Matrix Multiplication - *Undeclared*
    - 2.3: The Identity and Inverses - *CC BY 4.0*
    - 2.4: Finding the Inverse of a Matrix - *CC BY 4.0*
    - 2.5: Elementary Matrices - *CC BY 4.0*
    - 2.6: LU Factorization - *CC BY 4.0*
  - 3: Determinants - *CC BY 4.0*
    - 3.1: Basic Techniques - *CC BY 4.0*
      - 3.1E: Exercises for Section 3.1 - *CC BY 4.0*
    - 3.2: Properties of Determinants - *CC BY 4.0*
      - 3.2E: Exercises for Section 3.2 - *CC BY 4.0*
    - 3.3: Application of the Determinant to Inverses; Cramer's Rule - *CC BY 4.0*
      - 3.3E: Exercises for Section 3.3 - *CC BY 4.0*
    - 3.4: Determinants and Geometry - *CC BY 4.0*
      - 3.4E: Exercises for Section 3.4 - *CC BY 4.0*
  - 4: Vector Spaces - R^n - *CC BY 4.0*
    - 4.1: Subspaces - *GNU Free Documentation License*
    - 4.2: Review of Vectors - *CC BY 4.0*
      - 4.2E: Exercises for Section 4.1 - *CC BY 4.0*
    - 4.3: Dot and Cross Product - *CC BY 4.0*
      - 4.3E: Exercises for Section 4.2 - *CC BY 4.0*
    - 4.4: Lines and Planes - *CC BY 4.0*
      - 4.4E: Exercises for Section 4.3 - *CC BY 4.0*
    - 4.5: Spanning Sets in R^n - *CC BY 4.0*
      - 4.5.E: Exercise for Section 4.4 - *CC BY 4.0*
    - 4.6: Linear Independence - *CC BY 4.0*
      - 4.6.E: Exercise for Section 4.5 - *CC BY 4.0*
    - 4.7: Subspaces - *GNU Free Documentation License*
    - 4.8: Subspaces and Bases - *CC BY 4.0*
      - 4.8.E: Exercise for Section 4.6 - *CC BY 4.0*
    - 4.9: Row, Column and Null Spaces - *CC BY 4.0*
      - 4.9.E: Exercise for Section 4.7 - *CC BY 4.0*
    - 4.10: Dot Products and Orthogonality - *GNU Free Documentation License*
    - 4.11: Orthogonal Vectors and Matrices - *CC BY 4.0*
      - 4.11.E: Exercise for Section 4.8 - *CC BY 4.0*
    - 4.12: Gram-Schmidt Process - *CC BY 4.0*
      - 4.12.E: Exercises for Section 4.9 - *CC BY 4.0*
    - 4.13: Orthogonal Complements - *GNU Free Documentation License*
    - 4.14: Orthogonal Projections - *CC BY 4.0*
      - 4.14.E: Exercise for Section 4.10 - *CC BY 4.0*
    - 4.15: Orthogonal Projection - *GNU Free Documentation License*

- 4.16: Least Squares Approximation - *CC BY 4.0*
  - 4.16.E: Exercises for Section 4.11 - *CC BY 4.0*
- 4.17: Vector Spaces and Subspace - *Undeclared*
- 4.18: Amira Vector Space Axioms - *Undeclared*
- 4.19: Amira Vector Subspace of R^n - *Undeclared*
- 4.20: The Real Vector Space ^R^n - *Undeclared*
- 4.E: Exercises - *CC BY 4.0*

◦ 5: Linear Transformations - *GNU Free Documentation License*
- 5.1: Prelude to Linear Transformations and Matrix Algebra - *GNU Free Documentation License*
- 5.2: Linear Transformations - *GNU Free Documentation License*
- 5.3: Matrix Transformations - *GNU Free Documentation License*
- 5.4: One-to-one and Onto Transformations - *GNU Free Documentation License*
- 5.5: The Invertible Matrix Theorem - *GNU Free Documentation License*

◦ 6: Eigenvalues and Eigenvectors - *CC BY 4.0*
- 6.1: An introduction to Eigenvalues and Eigenvectors - *CC BY 4.0*
- 6.2: Finding eigenvalues and eigenvectors - *CC BY 4.0*
- 6.3: Diagonalization, similarity, and powers of a matrix - *CC BY 4.0*
- 6.4: Dynamical Systems - *CC BY 4.0*
- 6.5: Markov chains and Google's PageRank algorithm - *CC BY 4.0*

◦ 7: Honors Projects - *Undeclared*
- 7.1: Stochastic Matrices - *GNU Free Documentation License*
- 7.2: An Application to Markov Chains - *CC BY-NC-SA 4.0*
  - 7.2.1: Absorbing Markov Chains - *CC BY 4.0*
    - 7.2.1.1: Absorbing Markov Chains (Exercises) - *CC BY 4.0*
  - 7.2.2: Introduction to Markov Chains - *CC BY 4.0*
    - 7.2.2.1: Introduction to Markov Chains (Exercises) - *CC BY 4.0*
  - 7.2.3: Regular Markov Chains - *CC BY 4.0*
    - 7.2.3.1: Regular Markov Chains (Exercises) - *CC BY 4.0*
- 7.3: The Method of Least Squares - *GNU Free Documentation License*
- 7.4: An Application to Linear Recurrences - *CC BY-NC-SA 4.0*
  - 7.4E: An Application to Linear Recurrences Exercises - *Undeclared*
- 7.5: An Application to Systems of Differential Equations - *CC BY-NC-SA 4.0*
  - 7.5E: An Application to Systems of Differential Equations Exercises - *Undeclared*
- 7.6: An Application to Computer Graphics - *CC BY-NC-SA 4.0*
  - 7.6E: An Application to Computer Graphics Exercises - *Undeclared*
- 7.7: An Application to Input-Output Economic Models - *CC BY-NC-SA 4.0*
  - 7.7E: An Application to Input-Output Economic Models Exercises - *Undeclared*
- 7.8: The Complex Number System - *CC BY-NC-SA 3.0*
  - 7.8.1: Solve Systems of Linear Equations with Complex Coefficients - *Undeclared*

◦ Back Matter - *Undeclared*
- Index - *Undeclared*
- Glossary - *Undeclared*
- Detailed Licensing - *Undeclared*
- Detailed Licensing - *Undeclared*
- Detailed Licensing - *Undeclared*

This page titled Linear Algebra and Its Application is shared under a license and was authored, remixed, and/or curated by Amira Alkeswani, Cañada College.