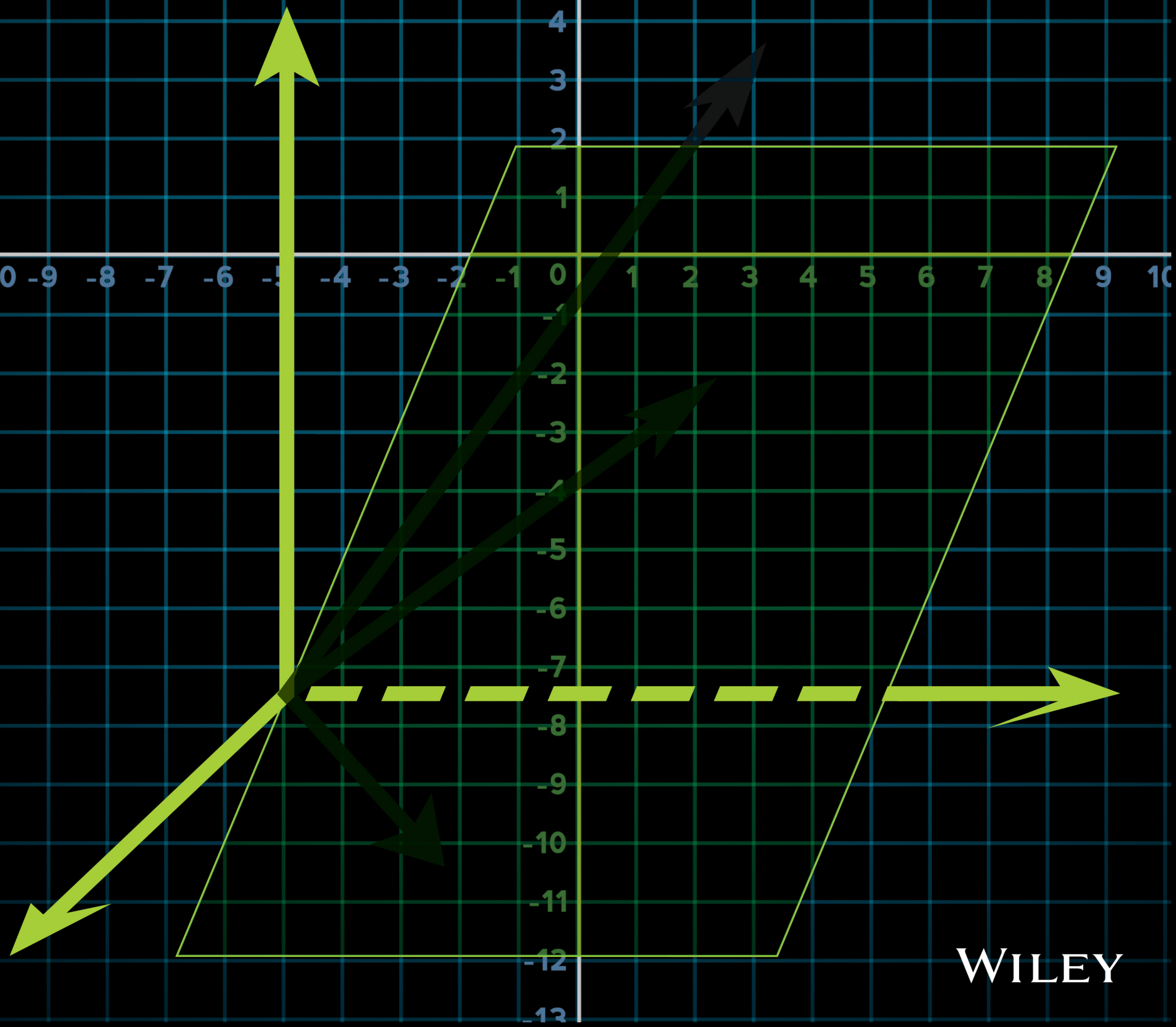
ELEMENTARY LINEAR ALGEBRA

HOWARD ANTON • ANTON KAUL

TWELFTH EDITION



WILEY

Applications and Historical Topics

Section in the Applications Version
****Web Module***

Section in the Applications Version
****Web Module***

Elementary Linear Algebra

12th Edition

Elementary Linear Algebra

12th Edition

HOWARD ANTON
Professor Emeritus, Drexel University

ANTON KAUL
Professor, California Polytechnic State University

WILEY

VICE PRESIDENT AND EDITORIAL DIRECTOR	Laurie Rosatone
EXECUTIVE EDITOR	Terri Ward
PRODUCT DESIGNER	Melissa Whelan
PRODUCT DESIGN MANAGER	Tom Kulesa
PRODUCT DESIGN EDITORORIAL ASSISTANT	Kimberly Eskin
EDITORIAL ASSISTANT	Crystal Franks
SENIOR CONTENT MANAGER	Valeri Zaborski
SENIOR PRODUCTION EDITOR	Laura Abrams
SENIOR MARKETING MANAGER	Michael MacDougald
PHOTO EDITOR	Billy Ray
COVER DESIGNER	Tom Nery
COVER AND CHAPTER OPENER PHOTO	© Lantica/Shutterstock

This book was set in STIXTwoText by MPS Limited and printed and bound by Quad Graphics/Versailles. The cover was printed by Quad Graphics/Versailles.

This book is printed on acid-free paper. ♾

Founded in 1807, John Wiley & Sons, Inc. has been a valued source of knowledge and understanding for more than 200 years, helping people around the world meet their needs and fulfill their aspirations. Our company is built on a foundation of principles that include responsibility to the communities we serve and where we live and work. In 2008, we launched a Corporate Citizenship Initiative, a global effort to address the environmental, social, economic, and ethical challenges we face in our business. Among the issues we are addressing are carbon impact, paper specifications and procurement, ethical conduct within our business and among our vendors, and community and charitable support. For more information, please visit our website: www.wiley.com/go/citizenship.

ISBN-13: 978-1-119-40677-8

The inside back cover will contain printing identification and country of origin if omitted from this page. In addition, if the ISBN on the back cover differs from the ISBN on this page, the one on the back cover is correct.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To

My wife, Pat

My children, Brian, David, and Lauren

My parents, Shirley and Benjamin

In memory of Prof. Leon Bahar,
who fostered my love of mathematics

My benefactor, Stephen Girard (1750–1831),
whose philanthropy changed my life

Howard Anton

To

My wife, Michelle, and my boys, Ulysses and Seth

Anton Kaul

About the Authors

HOWARD ANTON obtained his B.A. from Lehigh University, his M.A. from the University of Illinois, and his Ph.D. from the Polytechnic Institute of Brooklyn (now part of New York University), all in mathematics. In the early 1960s he was employed by the Burroughs Corporation at Cape Canaveral, Florida, where he worked on mathematical problems in the manned space program. In 1968 he joined the Mathematics Department of Drexel University, where he taught and did research until 1983. Since then he has devoted the majority of his time to textbook writing and activities for mathematical associations. Dr. Anton was president of the Eastern Pennsylvania and Delaware Section of the Mathematical Association of America, served on the Board of Governors of that organization, and guided the creation of its Student Chapters. He is the coauthor of a popular calculus text and has authored numerous research papers in functional analysis, topology, and approximation theory. His textbooks are among the most widely used in the world. There are now more than 200 versions of his books, including translations into Spanish, Arabic, Portuguese, Italian, Indonesian, French, and Japanese. For relaxation, Dr. Anton enjoys travel and photography. This text is the recipient of the *Textbook Excellence Award* by Textbook & Academic Authors Association.

ANTON KAUL received his B.S. from UC Davis and his M.S. and Ph.D. from Oregon State University. He held positions at the University of South Florida and Tufts University before joining the faculty at Cal Poly, San Luis Obispo in 2003, where he is currently a professor in the Mathematics Department. In addition to his work on mathematics textbooks, Dr. Kaul has done research in the area of geometric group theory and has published journal articles on Coxeter groups and their automorphisms. He is also an avid baseball fan and old-time banjo player.

Preface

We are proud that this book is the recipient of the *Textbook Excellence Award* from the Text & Academic Authors Association. Its quality owes much to the many professors who have taken the time to write and share their pedagogical expertise. We thank them all.

This 12th edition of *Elementary Linear Algebra* has a new contemporary design, many new exercises, and some organizational changes suggested by the classroom experience of many users. However, the fundamental philosophy of this book has not changed. It provides an introductory treatment of linear algebra that is suitable for a first undergraduate course. Its aim is to present the fundamentals of the subject in the clearest possible way, with sound pedagogy being the main consideration. Although calculus is not a prerequisite, some optional material here is clearly marked for students with a calculus background. If desired, that material can be omitted without loss of continuity. Technology is not required to use this text. However, clearly marked exercises that require technology are included for those who would like to use MATLAB, Mathematica, Maple, or other software with linear algebra capabilities. Supporting data files are posted on both of the following sites:

www.howardanton.com
www.wiley.com/college/anton

Summary of Changes in this Edition

Many parts of the text have been revised based on an extensive set of reviews. Here are the primary changes:

- **Earlier Linear Transformations** — Selected material on linear transformations that was covered later in the previous edition has been moved to Chapter 1 to provide a more complete early introduction to the topic. Specifically, some of the material in Sections 4.10 and 4.11 of the previous edition was extracted to form the new Section 1.9, and the remaining material is now in Section 8.6.
- **New Section 4.3 Devoted to Spanning Sets** — Section 4.2 of the previous edition dealt with both subspaces and spanning sets. Classroom experience has suggested that too many concepts were being introduced at once, so we have slowed down the pace and split off the material on spanning sets to create a new Section 4.3.
- **New Examples** — New examples have been added, where needed, to support the exercise sets.
- **New Exercises** — New exercises have been added with special attention to the expanded early introduction to linear transformations.

Alternative Version

As detailed on the front endpapers, this version of the text includes numerous real-world applications. However, instructors who want to cover a range of applications in more detail might consider the alternative version of this text, *Elementary Linear Algebra with Applications* by Howard Anton, Chris Rorres, and Anton Kaul (ISBN 978-1-119-40672-3). That version contains the first nine chapters of this text plus a tenth chapter with 20 detailed applications. Additional applications, listed in the Table of Contents, can be found on the the websites that accompany this text.

Hallmark Features

- **Interrelationships Among Concepts** — One of our main pedagogical goals is to convey to the student that linear algebra is not a collection of isolated definitions and techniques, but is rather a cohesive subject with interrelated ideas. One way in which we do this is by using a crescendo of theorems labeled "Equivalent Statements" that continually revisit relationships among systems of equations, matrices, determinants, vectors, linear transformations, and eigenvalues. To get a general sense of this pedagogical technique see Theorems 1.5.3, 1.6.4, 2.3.8, 4.9.8, 5.1.5, 6.4.5, and 8.2.4.
- **Smooth Transition to Abstraction** — Because the transition from Euclidean spaces to general vector spaces is difficult for many students, considerable effort is devoted to explaining the purpose of abstraction and helping the student to "visualize" abstract ideas by drawing analogies to familiar geometric ideas.
- **Mathematical Precision** — We try to be as mathematically precise as is reasonable for students at this level. But we recognize that mathematical precision is something to be learned, so proofs are presented in a patient style that is tailored for beginners.
- **Suitability for a Diverse Audience** — The text is designed to serve the needs of students in engineering, computer science, biology, physics, business, and economics, as well as those majoring in mathematics.
- **Historical Notes** — We feel that it is important to give students a sense of mathematical history and to convey that real people created the mathematical theorems and equations they are studying. Accordingly, we have included numerous "Historical Notes" that put various topics in historical perspective.

About the Exercises

- **Graded Exercise Sets** — Each exercise set begins with routine drill problems and progresses to problems with more substance. These are followed by three categories of problems, the first focusing on proofs, the second on true/false exercises, and the third on problems requiring technology. This compartmentalization is designed to simplify the instructor's task of selecting exercises for homework.
- **True/False Exercises** — The true/false exercises are designed to check conceptual understanding and logical reasoning. To avoid pure guesswork, the students are required to justify their responses in some way.
- **Proof Exercises** — Linear algebra courses vary widely in their emphasis on proofs, so exercises involving proofs have been grouped for easy identification. Appendix A provides students some guidance on proving theorems.
- **Technology Exercises** — Exercises that require technology have also been grouped. To avoid burdening the student with typing, the relevant data files have been posted on the websites that accompany this text.
- **Supplementary Exercises** — Each chapter ends with a set of exercises that draws from all the sections in the chapter.

Supplementary Materials for Students Available on the Web

- **Self Testing Review** — This edition also has an exciting new supplement, called the *Linear Algebra Flash-Card Review*. It is a self-study testing system based on the SQ3R study method that students can use to check their mastery of virtually every fundamental concept in this text. It is integrated into WileyPlus, and is available as a free app for iPads. The app can be obtained from the Apple Store by searching for:

  Anton Linear Algebra FlashCard Review

- **Student Solutions Manual** — This supplement provides detailed solutions to most odd-numbered exercises.
- **Maple Data Files** — Data files in Maple format for the technology exercises that are posted on the websites that accompany this text.
- **Mathematica Data Files** — Data files in Mathematica format for the technology exercises that are posted on the websites that accompany this text.
- **MATLAB Data Files** — Data files in MATLAB format for the technology exercises that are posted on the websites that accompany this text.
- **CSV Data Files** — Data files in CSV format for the technology exercises that are posted on the websites that accompany this text.
- **How to Read and Do Proofs** — A series of videos created by Prof. Daniel Solow of the Weatherhead School of Management, Case Western Reserve University, that present various strategies for proving theorems. These are available through WileyPLUS as well as the websites that accompany this text. There is also a guide for locating the appropriate videos for specific proofs in the text.
- **MATLAB Linear Algebra Manual and Laboratory Projects** — This supplement contains a set of laboratory projects written by Prof. Dan Seth of West Texas A&M University. It is designed to help students learn key linear algebra concepts by using MATLAB and is available in PDF form without charge to students at schools adopting the 12th edition of this text.
- **Data Files** — The data files needed for the MATLAB Linear Algebra Manual and Lab Projects supplement.
- **How to Open and Use MATLAB Files** — Instructional document on how to download, open, and use the MATLAB files accompanying this text.

Supplementary Materials for Instructors

- **Instructor Solutions Manual** — This supplement provides worked-out solutions to most exercises in the text.
- **PowerPoint Slides** — A series of slides that display important definitions, examples, graphics, and theorems in the book. These can also be distributed to students as review materials or to simplify note-taking.
- **Test Bank** — Test questions and sample examinations in PDF or LaTeX form.
- **Image Gallery** — Digital repository of images from the text that instructors may use to generate their own PowerPoint slides.
- **WileyPLUS** — An online environment for effective teaching and learning. WileyPLUS builds student confidence by taking the guesswork out of studying and by providing a clear roadmap of what to do, how to do it, and whether it was done right. Its purpose is to motivate and foster initiative so instructors can have a greater impact on classroom achievement and beyond.
- **WileyPLUS Question Index** — This document lists every question in the current WileyPLUS course and provides the name, associated learning objective, question type, and difficulty level for each. If available, it also shows the correlation between the previous edition WileyPLUS question and the current WileyPLUS question, so instructors can conveniently see the evolution of a question and reuse it from previous semester assignments.

A Guide for the Instructor

Although linear algebra courses vary widely in content and philosophy, most courses fall into two categories, those with roughly 40 lectures, and those with roughly 30 lectures. Accordingly, we have created the following long and short templates as possible starting points for constructing your own course outline. Keep in mind that these are just guides, and we fully expect that you will want to customize them to fit your own interests and requirements. Neither of these sample templates includes applications, so keep that in mind as you work with them.

	Long Template	Short Template
Chapter 1: Systems of Linear Equations and Matrices	8 lectures	6 lectures
Chapter 2: Determinants	3 lectures	3 lectures
Chapter 3: Euclidean Vector Spaces	4 lectures	3 lectures
Chapter 4: General Vector Spaces	8 lectures	7 lectures
Chapter 5: Eigenvalues and Eigenvectors	3 lectures	3 lectures
Chapter 6: Inner Product Spaces	3 lectures	2 lectures
Chapter 7: Diagonalization and Quadratic Forms	4 lectures	3 lectures
Chapter 8: General Linear Transformations	4 lectures	2 lectures
Chapter 9: Numerical Methods	2 lectures	1 lecture
Total:	**39 lectures**	**30 lectures**

Reviewers

The following people reviewed the plans for this edition, critiqued much of the content, and provided insightful pedagogical advice:

Charles Ekene Chika, *University of Texas at Dallas*

Marian Hukle, *University of Kansas*

Bin Jiang, *Portland State University*

Mike Panahi, *El Centro College*

Christopher Rasmussen, *Wesleyan University*

Nathan Reff, *The College at Brockport: SUNY*

Mark Smith, *Miami University*

Rebecca Swanson, *Colorado School of Mines*

R. Scott Williams, *University of Central Oklahoma*

Pablo Zafra, *Kean University*

Special Contributions

Our deep appreciation is due to a number of people who have contributed to this edition in many ways:

Prof. Mark Smith, who critiqued the FlashCard program and suggested valuable improvements to the text exposition.

Prof. Derek Hein, whose keen eye helped us to correct some subtle inaccuracies.

Susan Raley, who coordinated the production process and whose attention to detail made a very complex project run smoothly.

Prof. Roger Lipsett, whose mathematical expertise and detailed review of the manuscript has contributed greatly to its accuracy.

The Wiley Team, Laurie Rosatone, Terri Ward, Melissa Whelan, Tom Kulesa, Kimberly Eskin, Crystal Franks, Laura Abrams, Billy Ray, and Tom Nery each of whom contributed their experience, skill, and expertise to the project.

HOWARD ANTON
ANTON KAUL

Contents

CHAPTER 1

Systems of Linear Equations and Matrices

CHAPTER CONTENTS

Introduction

Information in science, business, and mathematics is often organized into rows and columns to form rectangular arrays called "matrices" (plural of "matrix"). Matrices often appear as tables of numerical data that arise from physical observations, but they occur in various mathematical contexts as well. For example, we will see in this chapter that all of the information required to solve a system of equations such as

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

is embodied in the matrix

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

and that the solution of the system can be obtained by performing appropriate operations on this matrix. This is particularly important in developing computer programs for solving systems of equations because computers are well suited for manipulating arrays of numerical information. However, matrices are not simply a notational tool for solving systems of equations; they can be viewed as mathematical objects in their own right, and there is a rich and important theory associated with them that has a multitude of practical applications. It is the study of matrices and related topics that forms the mathematical field that we call "linear algebra." In this chapter we will begin our study of matrices.

1.1 Introduction to Systems of Linear Equations

Systems of linear equations and their solutions constitute one of the major topics that we will study in this course. In this first section we will introduce some basic terminology and discuss a method for solving such systems.

Linear Equations

Recall that in two dimensions a line in a rectangular xy-coordinate system can be represented by an equation of the form

$$ax + by = c \quad (a, b \text{ not both } 0)$$

and in three dimensions a plane in a rectangular xyz-coordinate system can be represented by an equation of the form

$$ax + by + cz = d \quad (a, b, c \text{ not all } 0)$$

These are examples of "linear equations," the first being a linear equation in the variables x and y and the second a linear equation in the variables x, y, and z. More generally, we define a ***linear equation*** in the n variables $x_1, x_2, \dots, x_n$ to be one that can be expressed in the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b \tag{1}$$

where $a_1, a_2, \dots, a_n$ and b are constants, and the a's are not all zero. In the special cases where $n = 2$ or $n = 3$, we will often use variables without subscripts and write linear equations as

$$a_1x + a_2y = b \tag{2}$$

$$a_1x + a_2y + a_3z = b \tag{3}$$

In the special case where $b = 0$, Equation (1) has the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0 \tag{4}$$

which is called a ***homogeneous linear equation*** in the variables $x_1, x_2, \dots, x_n$.

EXAMPLE 1 | Linear Equations

Observe that a linear equation does not involve any products or roots of variables. All variables occur only to the first power and do not appear, for example, as arguments of trigonometric, logarithmic, or exponential functions. The following are linear equations:

$$x + 3y = 7 \qquad x_1 - 2x_2 - 3x_3 + x_4 = 0$$

$$\tfrac{1}{2}x - y + 3z = -1 \qquad x_1 + x_2 + \cdots + x_n = 1$$

The following are not linear equations:

$$x + 3y^2 = 4 \qquad 3x + 2y - xy = 5$$

$$\sin x + y = 0 \qquad \sqrt{x_1} + 2x_2 + x_3 = 1$$

A finite set of linear equations is called a ***system of linear equations*** or, more briefly, a ***linear system***. The variables are called ***unknowns***. For example, system (5) that follows has unknowns x and y, and system (6) has unknowns x_1, x_2, and x_3.

$$\begin{array}{ll} 5x + y = 3 & 4x_1 - x_2 + 3x_3 = -1 \\ 2x - y = 4 & 3x_1 + x_2 + 9x_3 = -4 \end{array} \tag{5–6}$$

A general linear system of m equations in the n unknowns $x_1, x_2, \dots, x_n$ can be written as

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array} \tag{7}$$

The double subscripting on the coefficients a_{ij} of the unknowns gives their location in the system—the first subscript indicates the equation in which the coefficient occurs, and the second indicates which unknown it multiplies. Thus, a_{12} is in the first equation and multiplies x_2.

A ***solution*** of a linear system in n unknowns $x_1, x_2, \dots, x_n$ is a sequence of n numbers $s_1, s_2, \dots, s_n$ for which the substitution

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

makes each equation a true statement. For example, the system in (5) has the solution

$$x = 1, \quad y = -2$$

and the system in (6) has the solution

$$x_1 = 1, \quad x_2 = 2, \quad x_3 = -1$$

These solutions can be written more succinctly as

$$(1, -2) \quad \text{and} \quad (1, 2, -1)$$

in which the names of the variables are omitted. This notation allows us to interpret these solutions geometrically as points in two-dimensional and three-dimensional space. More generally, a solution

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

of a linear system in n unknowns can be written as

$$(s_1, s_2, \dots, s_n)$$

which is called an ***ordered n-tuple***. With this notation it is understood that all variables appear in the same order in each equation. If $n = 2$, then the n-tuple is called an ***ordered pair***, and if $n = 3$, then it is called an ***ordered triple***.

Linear Systems in Two and Three Unknowns

Linear systems in two unknowns arise in connection with intersections of lines. For example, consider the linear system

$$\begin{aligned} a_1x + b_1y &= c_1 \\ a_2x + b_2y &= c_2 \end{aligned}$$

in which the graphs of the equations are lines in the xy-plane. Each solution (x, y) of this system corresponds to a point of intersection of the lines, so there are three possibilities (**Figure 1.1.1**):

1. The lines may be parallel and distinct, in which case there is no intersection and consequently no solution.
2. The lines may intersect at only one point, in which case the system has exactly one solution.
3. The lines may coincide, in which case there are infinitely many points of intersection (the points on the common line) and consequently infinitely many solutions.

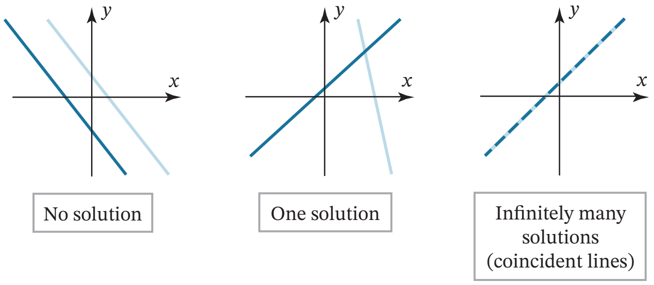


FIGURE 1.1.1

In general, we say that a linear system is ***consistent*** if it has at least one solution and ***inconsistent*** if it has no solutions. Thus, a *consistent* linear system of two equations in two unknowns has either one solution or infinitely many solutions—there are no other possibilities. The same is true for a linear system of three equations in three unknowns

$$
\begin{aligned}
a_1x + b_1y + c_1z &= d_1 \\
a_2x + b_2y + c_2z &= d_2 \\
a_3x + b_3y + c_3z &= d_3
\end{aligned}
$$

in which the graphs of the equations are planes. The solutions of the system, if any, correspond to points where all three planes intersect, so again we see that there are only three possibilities—no solutions, one solution, or infinitely many solutions (**Figure 1.1.2**).

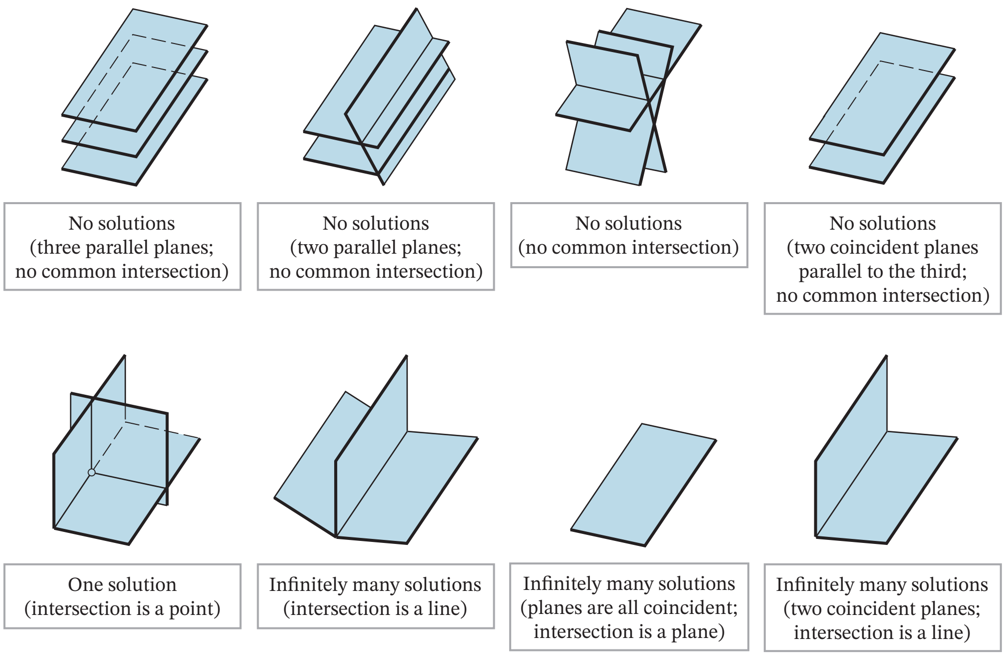


FIGURE 1.1.2

We will prove later that our observations about the number of solutions of linear systems of two equations in two unknowns and linear systems of three equations in three unknowns actually hold for *all* linear systems. That is:

> *Every system of linear equations has zero, one, or infinitely many solutions. There are no other possibilities.*

EXAMPLE 2 | A Linear System with One Solution

Solve the linear system

$$\begin{aligned} x - y &= 1 \\ 2x + y &= 6 \end{aligned}$$

Solution We can eliminate x from the second equation by adding -2 times the first equation to the second. This yields the simplified system

$$\begin{aligned} x - y &= 1 \\ 3y &= 4 \end{aligned}$$

From the second equation we obtain $y = \frac{4}{3}$, and on substituting this value in the first equation we obtain $x = 1 + y = \frac{7}{3}$. Thus, the system has the unique solution

$$x = \tfrac{7}{3}, \quad y = \tfrac{4}{3}$$

Geometrically, this means that the lines represented by the equations in the system intersect at the single point $\left(\frac{7}{3}, \frac{4}{3}\right)$. We leave it for you to check this by graphing the lines.

EXAMPLE 3 | A Linear System with No Solutions

Solve the linear system

$$\begin{aligned} x + y &= 4 \\ 3x + 3y &= 6 \end{aligned}$$

Solution We can eliminate x from the second equation by adding -3 times the first equation to the second equation. This yields the simplified system

$$\begin{aligned} x + y &= 4 \\ 0 &= -6 \end{aligned}$$

The second equation is contradictory, so the given system has no solution. Geometrically, this means that the lines corresponding to the equations in the original system are parallel and distinct. We leave it for you to check this by graphing the lines or by showing that they have the same slope but different y-intercepts.

EXAMPLE 4 | A Linear System with Infinitely Many Solutions

Solve the linear system

$$\begin{aligned} 4x - 2y &= 1 \\ 16x - 8y &= 4 \end{aligned}$$

Solution We can eliminate x from the second equation by adding -4 times the first equation to the second. This yields the simplified system

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

The second equation does not impose any restrictions on x and y and hence can be omitted. Thus, the solutions of the system are those values of x and y that satisfy the single equation

$$4x - 2y = 1 \tag{8}$$

Geometrically, this means the lines corresponding to the two equations in the original system coincide. One way to describe the solution set is to solve this equation for x in terms of y to

In Example 4 we could have also obtained parametric equations for the solutions by solving (8) for y in terms of x and letting $x = t$ be the parameter. The resulting parametric equations would look different but would define the same solution set.

obtain $x = \frac{1}{4} + \frac{1}{2}y$ and then assign an arbitrary value t (called a ***parameter***) to y. This allows us to express the solution by the pair of equations (called ***parametric equations***)

$$x = \tfrac{1}{4} + \tfrac{1}{2}t, \quad y = t$$

We can obtain specific numerical solutions from these equations by substituting numerical values for the parameter t. For example, $t = 0$ yields the solution $\left(\frac{1}{4}, 0\right)$, $t = 1$ yields the solution $\left(\frac{3}{4}, 1\right)$, and $t = -1$ yields the solution $\left(-\frac{1}{4}, -1\right)$. You can confirm that these are solutions by substituting their coordinates into the given equations.

EXAMPLE 5 | A Linear System with Infinitely Many Solutions

Solve the linear system

$$\begin{aligned} x - y + 2z &= 5 \\ 2x - 2y + 4z &= 10 \\ 3x - 3y + 6z &= 15 \end{aligned}$$

Solution This system can be solved by inspection, since the second and third equations are multiples of the first. Geometrically, this means that the three planes coincide and that those values of x, y, and z that satisfy the equation

$$x - y + 2z = 5 \tag{9}$$

automatically satisfy all three equations. Thus, it suffices to find the solutions of (9). We can do this by first solving this equation for x in terms of y and z, then assigning arbitrary values r and s (parameters) to these two variables, and then expressing the solution by the three parametric equations

$$x = 5 + r - 2s, \quad y = r, \quad z = s$$

Specific solutions can be obtained by choosing numerical values for the parameters r and s. For example, taking $r = 1$ and $s = 0$ yields the solution $(6, 1, 0)$.

Augmented Matrices and Elementary Row Operations

As the number of equations and unknowns in a linear system increases, so does the complexity of the algebra involved in finding solutions. The required computations can be made more manageable by simplifying notation and standardizing procedures. For example, by mentally keeping track of the location of the $+$'s, the x's, and the $=$'s in the linear system

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

we can abbreviate the system by writing only the rectangular array of numbers

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

As noted in the introduction to this chapter, the term "matrix" is used in mathematics to denote a rectangular array of numbers. In a later section we will study matrices in detail, but for now we will only be concerned with augmented matrices for linear systems.

This is called the ***augmented matrix*** for the system. For example, the augmented matrix for the system of equations

$$\begin{aligned} x_1 + x_2 + 2x_3 &= 9 \\ 2x_1 + 4x_2 - 3x_3 &= 1 \\ 3x_1 + 6x_2 - 5x_3 &= 0 \end{aligned} \qquad \text{is} \qquad \begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{bmatrix}$$

Historical Note

Maxime Bôcher (1867–1918)

The first known use of augmented matrices appeared between 200 B.C. and 100 B.C. in a Chinese manuscript entitled *Nine Chapters of Mathematical Art*. The coefficients were arranged in columns rather than in rows, as today, but remarkably the system was solved by performing a succession of operations on the columns. The actual use of the term *augmented matrix* appears to have been introduced by the American mathematician Maxime Bôcher in his book *Introduction to Higher Algebra*, published in 1907. In addition to being an outstanding research mathematician and an expert in Latin, chemistry, philosophy, zoology, geography, meteorology, art, and music, Bôcher was an outstanding expositor of mathematics whose elementary textbooks were greatly appreciated by students and are still in demand today.

[Image: HUP Bocher, Maxime (1), olvwork650836]

The basic method for solving a linear system is to perform algebraic operations on the system that do not alter the solution set and that produce a succession of increasingly simpler systems, until a point is reached where it can be ascertained whether the system is consistent, and if so, what its solutions are. Typically, the algebraic operations are:

1. Multiply an equation through by a nonzero constant.
2. Interchange two equations.
3. Add a constant times one equation to another.

Since the rows (horizontal lines) of an augmented matrix correspond to the equations in the associated system, these three operations correspond to the following operations on the rows of the augmented matrix:

1. Multiply a row through by a nonzero constant.
2. Interchange two rows.
3. Add a constant times one row to another.

These are called ***elementary row operations*** on a matrix.

In the following example we will illustrate how to use elementary row operations and an augmented matrix to solve a linear system in three unknowns. Since a systematic procedure for solving linear systems will be developed in the next section, do not worry about how the steps in the example were chosen. Your objective here should be simply to understand the computations.

EXAMPLE 6 | Using Elementary Row Operations

In the left column we solve a system of linear equations by operating on the equations in the system, and in the right column we solve the same system by operating on the rows of the augmented matrix.

$$\begin{aligned} x + y + 2z &= 9 \\ 2x + 4y - 3z &= 1 \\ 3x + 6y - 5z &= 0 \end{aligned} \qquad \begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{bmatrix}$$

Add -2 times the first equation to the second to obtain

$$\begin{aligned} x + y + 2z &= 9 \\ 2y - 7z &= -17 \\ 3x + 6y - 5z &= 0 \end{aligned}$$

Add -2 times the first row to the second to obtain

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 3 & 6 & -5 & 0 \end{bmatrix}$$

Add -3 times the first equation to the third to obtain

$$\begin{aligned} x + y + 2z &= 9 \\ 2y - 7z &= -17 \\ 3y - 11z &= -27 \end{aligned}$$

Add -3 times the first row to the third to obtain

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 0 & 3 & -11 & -27 \end{bmatrix}$$

Multiply the second equation by $\frac{1}{2}$ to obtain

$$\begin{aligned} x + y + 2z &= 9 \\ y - \tfrac{7}{2}z &= -\tfrac{17}{2} \\ 3y - 11z &= -27 \end{aligned}$$

Multiply the second row by $\frac{1}{2}$ to obtain

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 3 & -11 & -27 \end{bmatrix}$$

Add -3 times the second equation to the third to obtain

$$\begin{aligned} x + y + 2z &= 9 \\ y - \tfrac{7}{2}z &= -\tfrac{17}{2} \\ -\tfrac{1}{2}z &= -\tfrac{3}{2} \end{aligned}$$

Add -3 times the second row to the third to obtain

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & -\frac{1}{2} & -\frac{3}{2} \end{bmatrix}$$

Multiply the third equation by -2 to obtain

$$\begin{aligned} x + y + 2z &= 9 \\ y - \tfrac{7}{2}z &= -\tfrac{17}{2} \\ z &= 3 \end{aligned}$$

Multiply the third row by -2 to obtain

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

Add -1 times the second equation to the first to obtain

$$\begin{aligned} x + \tfrac{11}{2}z &= \tfrac{35}{2} \\ y - \tfrac{7}{2}z &= -\tfrac{17}{2} \\ z &= 3 \end{aligned}$$

Add -1 times the second row to the first to obtain

$$\begin{bmatrix} 1 & 0 & \frac{11}{2} & \frac{35}{2} \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

Add $-\frac{11}{2}$ times the third equation to the first and $\frac{7}{2}$ times the third equation to the second to obtain

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

Add $-\frac{11}{2}$ times the third row to the first and $\frac{7}{2}$ times the third row to the second to obtain

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

The solution $x = 1,\ y = 2,\ z = 3$ is now evident.

The solution in this example can also be expressed as the ordered triple $(1, 2, 3)$ with the understanding that the numbers in the triple are in the same order as the variables in the system, namely, x, y, z.

Exercise Set 1.1

1. In each part, determine whether the equation is linear in x_1, x_2, and x_3.

a. $x_1 + 5x_2 - \sqrt{2}x_3 = 1$ **b.** $x_1 + 3x_2 + x_1x_3 = 2$

c. $x_1 = -7x_2 + 3x_3$ **d.** $x_1^{-2} + x_2 + 8x_3 = 5$

e. $x_1^{3/5} - 2x_2 + x_3 = 4$ **f.** $\pi x_1 - \sqrt{2}x_2 = 7^{1/3}$

2. In each part, determine whether the equation is linear in x and y.

a. $2^{1/3}x + \sqrt{3}y = 1$ **b.** $2x^{1/3} + 3\sqrt{y} = 1$

c. $\cos\left(\frac{\pi}{7}\right)x - 4y = \log 3$ **d.** $\frac{\pi}{7}\cos x - 4y = 0$

e. $xy = 1$ **f.** $y + 7 = x$

3. Using the notation of Formula (7), write down a general linear system of

a. two equations in two unknowns.

b. three equations in three unknowns.

c. two equations in four unknowns.

4. Write down the augmented matrix for each of the linear systems in Exercise 3.

In each part of Exercises 5–6, find a system of linear equations in the unknowns $x_1, x_2, x_3, \ldots$, that corresponds to the given augmented matrix.

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

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

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

In each part of Exercises 7–8, find the augmented matrix for the linear system.

7. a. $\begin{aligned} -2x_1 &= 6 \\ 3x_1 &= 8 \\ 9x_1 &= -3 \end{aligned}$ **b.** $\begin{aligned} 6x_1 - x_2 + 3x_3 &= 4 \\ 5x_2 - x_3 &= 1 \end{aligned}$

c. $\begin{aligned} 2x_2 - 3x_4 + x_5 &= 0 \\ -3x_1 - x_2 + x_3 &= -1 \\ 6x_1 + 2x_2 - x_3 + 2x_4 - 3x_5 &= 6 \end{aligned}$

8. a. $\begin{aligned} 3x_1 - 2x_2 &= -1 \\ 4x_1 + 5x_2 &= 3 \\ 7x_1 + 3x_2 &= 2 \end{aligned}$ **b.** $\begin{aligned} 2x_1 + 2x_3 &= 1 \\ 3x_1 - x_2 + 4x_3 &= 7 \\ 6x_1 + x_2 - x_3 &= 0 \end{aligned}$

c. $\begin{aligned} x_1 &= 1 \\ x_2 &= 2 \\ x_3 &= 3 \end{aligned}$

9. In each part, determine whether the given 3-tuple is a solution of the linear system

$$\begin{aligned} 2x_1 - 4x_2 - x_3 &= 1 \\ x_1 - 3x_2 + x_3 &= 1 \\ 3x_1 - 5x_2 - 3x_3 &= 1 \end{aligned}$$

a. $(3, 1, 1)$ **b.** $(3, -1, 1)$ **c.** $(13, 5, 2)$

d. $\left(\frac{13}{2}, \frac{5}{2}, 2\right)$ **e.** $(17, 7, 5)$

10. In each part, determine whether the given 3-tuple is a solution of the linear system

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

a. $\left(\frac{5}{7}, \frac{8}{7}, 1\right)$ **b.** $\left(\frac{5}{7}, \frac{8}{7}, 0\right)$ **c.** $(5, 8, 1)$

d. $\left(\frac{5}{7}, \frac{10}{7}, \frac{2}{7}\right)$ **e.** $\left(\frac{5}{7}, \frac{22}{7}, 2\right)$

11. In each part, solve the linear system, if possible, and use the result to determine whether the lines represented by the equations in the system have zero, one, or infinitely many points of intersection. If there is a single point of intersection, give its coordinates, and if there are infinitely many, find parametric equations for them.

a. $\begin{aligned} 3x - 2y &= 4 \\ 6x - 4y &= 9 \end{aligned}$ **b.** $\begin{aligned} 2x - 4y &= 1 \\ 4x - 8y &= 2 \end{aligned}$ **c.** $\begin{aligned} x - 2y &= 0 \\ x - 4y &= 8 \end{aligned}$

12. Under what conditions on a and b will the linear system have no solutions, one solution, infinitely many solutions?

$$\begin{aligned} 2x - 3y &= a \\ 4x - 6y &= b \end{aligned}$$

In each part of Exercises 13–14, use parametric equations to describe the solution set of the linear equation.

13. a. $7x - 5y = 3$

b. $3x_1 - 5x_2 + 4x_3 = 7$

c. $-8x_1 + 2x_2 - 5x_3 + 6x_4 = 1$

d. $3v - 8w + 2x - y + 4z = 0$

14. a. $x + 10y = 2$

b. $x_1 + 3x_2 - 12x_3 = 3$

c. $4x_1 + 2x_2 + 3x_3 + x_4 = 20$

d. $v + w + x - 5y + 7z = 0$

In Exercises 15–16, each linear system has infinitely many solutions. Use parametric equations to describe its solution set.

15. a. $\begin{aligned} 2x - 3y &= 1 \\ 6x - 9y &= 3 \end{aligned}$

b. $\begin{aligned} x_1 + 3x_2 - x_3 &= -4 \\ 3x_1 + 9x_2 - 3x_3 &= -12 \\ -x_1 - 3x_2 + x_3 &= 4 \end{aligned}$

16. a. $\begin{aligned} 6x_1 + 2x_2 &= -8 \\ 3x_1 + x_2 &= -4 \end{aligned}$ **b.** $\begin{aligned} 2x - y + 2z &= -4 \\ 6x - 3y + 6z &= -12 \\ -4x + 2y - 4z &= 8 \end{aligned}$

In Exercises 17–18, find a single elementary row operation that will create a 1 in the upper left corner of the given augmented matrix and will not create any fractions in its first row.

17. a. $\begin{bmatrix} -3 & -1 & 2 & 4 \\ 2 & -3 & 3 & 2 \\ 0 & 2 & -3 & 1 \end{bmatrix}$ **b.** $\begin{bmatrix} 0 & -1 & -5 & 0 \\ 2 & -9 & 3 & 2 \\ 1 & 4 & -3 & 3 \end{bmatrix}$

18. a. $\begin{bmatrix} 2 & 4 & -6 & 8 \\ 7 & 1 & 4 & 3 \\ -5 & 4 & 2 & 7 \end{bmatrix}$ **b.** $\begin{bmatrix} 7 & -4 & -2 & 2 \\ 3 & -1 & 8 & 1 \\ -6 & 3 & -1 & 4 \end{bmatrix}$

In Exercises 19–20, find all values of k for which the given augmented matrix corresponds to a consistent linear system.

19. a. $\begin{bmatrix} 1 & k & -4 \\ 4 & 8 & 2 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & k & -1 \\ 4 & 8 & -4 \end{bmatrix}$

20. a. $\begin{bmatrix} 3 & -4 & k \\ -6 & 8 & 5 \end{bmatrix}$ **b.** $\begin{bmatrix} k & 1 & -2 \\ 4 & -1 & 2 \end{bmatrix}$

21. The curve $y = ax^2 + bx + c$ shown in the accompanying figure passes through the points (x_1, y_1), (x_2, y_2), and (x_3, y_3). Show that the coefficients a, b, and c form a solution of the system of linear equations whose augmented matrix is

$$\begin{bmatrix} x_1^2 & x_1 & 1 & y_1 \\ x_2^2 & x_2 & 1 & y_2 \\ x_3^2 & x_3 & 1 & y_3 \end{bmatrix}$$

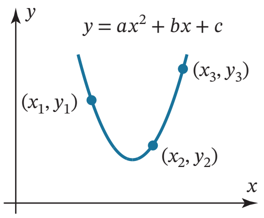


FIGURE Ex-21

22. Explain why each of the three elementary row operations does not affect the solution set of a linear system.

23. Show that if the linear equations

$$x_1 + kx_2 = c \qquad \text{and} \qquad x_1 + lx_2 = d$$

have the same solution set, then the two equations are identical (i.e., $k = l$ and $c = d$).

24. Consider the system of equations

$$\begin{aligned} ax + by &= k \\ cx + dy &= l \\ ex + fy &= m \end{aligned}$$

Discuss the relative positions of the lines $ax + by = k$, $cx + dy = l$, and $ex + fy = m$ when

a. the system has no solutions.

b. the system has exactly one solution.

c. the system has infinitely many solutions.

25. Suppose that a certain diet calls for 7 units of fat, 9 units of protein, and 16 units of carbohydrates for the main meal, and suppose that an individual has three possible foods to choose from to meet these requirements:

Food 1: Each ounce contains 2 units of fat, 2 units of protein, and 4 units of carbohydrates.

Food 2: Each ounce contains 3 units of fat, 1 unit of protein, and 2 units of carbohydrates.

Food 3: Each ounce contains 1 unit of fat, 3 units of protein, and 5 units of carbohydrates.

Let x, y, and z denote the number of ounces of the first, second, and third foods that the dieter will consume at the main meal. Find (but do not solve) a linear system in x, y, and z whose solution tells how many ounces of each food must be consumed to meet the diet requirements.

26. Suppose that you want to find values for a, b, and c such that the parabola $y = ax^2 + bx + c$ passes through the points $(1, 1)$, $(2, 4)$, and $(-1, 1)$. Find (but do not solve) a system of linear equations whose solutions provide values for a, b, and c. How many solutions would you expect this system of equations to have, and why?

27. Suppose you are asked to find three real numbers such that the sum of the numbers is 12, the sum of two times the first plus the second plus two times the third is 5, and the third number is one more than the first. Find (but do not solve) a linear system whose equations describe the three conditions.

True-False Exercises

TF. In parts **(a)–(h)** determine whether the statement is true or false, and justify your answer.

a. A linear system whose equations are all homogeneous must be consistent.

b. Multiplying a row of an augmented matrix through by zero is an acceptable elementary row operation.

c. The linear system

$$\begin{aligned} x - y &= 3 \\ 2x - 2y &= k \end{aligned}$$

cannot have a unique solution, regardless of the value of k.

d. A single linear equation with two or more unknowns must have infinitely many solutions.

e. If the number of equations in a linear system exceeds the number of unknowns, then the system must be inconsistent.

f. If each equation in a consistent linear system is multiplied through by a constant c, then all solutions to the new system can be obtained by multiplying solutions from the original system by c.

g. Elementary row operations permit one row of an augmented matrix to be subtracted from another.

h. The linear system with corresponding augmented matrix

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

is consistent.

Working with Technology

T1. Solve the linear systems in Examples 2, 3, and 4 to see how your technology utility handles the three types of systems.

T2. Use the result in Exercise 21 to find values of a, b, and c for which the curve $y = ax^2 + bx + c$ passes through the points $(-1, 1, 4)$, $(0, 0, 8)$, and $(1, 1, 7)$.

1.2 Gaussian Elimination

In this section we will develop a systematic procedure for solving systems of linear equations. The procedure is based on the idea of performing certain operations on the rows of the augmented matrix that simplify it to a form from which the solution of the system can be ascertained by inspection.

Considerations in Solving Linear Systems

When considering methods for solving systems of linear equations, it is important to distinguish between large systems that must be solved by computer and small systems that can be solved by hand. For example, there are many applications that lead to linear systems in thousands or even millions of unknowns. Large systems require special techniques to deal with issues of memory size, roundoff errors, solution time, and so forth. Such techniques are studied in the field of ***numerical analysis*** and will only be touched on in this text. However, almost all of the methods that are used for large systems are based on the ideas that we will develop in this section.

Echelon Forms

In Example 6 of the last section, we solved a linear system in the unknowns x, y, and z by reducing the augmented matrix to the form

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

from which the solution $x = 1$, $y = 2$, $z = 3$ became evident. This is an example of a matrix that is in ***reduced row echelon form***. To be of this form, a matrix must have the following properties:

1. If a row does not consist entirely of zeros, then the first nonzero number in the row is a 1. We call this a ***leading 1***.
2. If there are any rows that consist entirely of zeros, then they are grouped together at the bottom of the matrix.
3. In any two successive rows that do not consist entirely of zeros, the leading 1 in the lower row occurs farther to the right than the leading 1 in the higher row.
4. Each column that contains a leading 1 has zeros everywhere else in that column.

A matrix that has the first three properties is said to be in ***row echelon form***. (Thus, a matrix in reduced row echelon form is of necessity in row echelon form, but not conversely.)

EXAMPLE 1 | Row Echelon and Reduced Row Echelon Form

The following matrices are in reduced row echelon form.

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

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

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

EXAMPLE 2 | More on Row Echelon and Reduced Row Echelon Form

As Example 1 illustrates, a matrix in row echelon form has zeros below each leading 1, whereas a matrix in reduced row echelon form has zeros below *and above* each leading 1. Thus, with any real numbers substituted for the $*$'s, all matrices of the following types are in row echelon form:

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

All matrices of the following types are in reduced row echelon form:

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

If, by a sequence of elementary row operations, the augmented matrix for a system of linear equations is put in *reduced* row echelon form, then the solution set can be obtained either by inspection or by converting certain linear equations to parametric form. Here are some examples.

EXAMPLE 3 | Unique Solution

Suppose that the augmented matrix for a linear system in the unknowns x_1, x_2, x_3, and x_4 has been reduced by elementary row operations to

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

This matrix is in reduced row echelon form and corresponds to the equations

$$\begin{aligned} x_1 \qquad\qquad\qquad &= 3 \\ x_2 \qquad\qquad &= -1 \\ x_3 \qquad &= 0 \\ x_4 &= 5 \end{aligned}$$

Thus, the system has a unique solution, namely, $x_1 = 3, x_2 = -1, x_3 = 0, x_4 = 5$, which can also be expressed as the 4-tuple $(3, -1, 0, 5)$.

EXAMPLE 4 | Linear Systems in Three Unknowns

In each part, suppose that the augmented matrix for a linear system in the unknowns x, y, and z has been reduced by elementary row operations to the given reduced row echelon form. Solve the system.

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

Solution (*a*) The equation that corresponds to the last row of the augmented matrix is

$$0x + 0y + 0z = 1$$

Since this equation is not satisfied by any values of x, y, and z, the system is inconsistent.

Solution (*b*) The equation that corresponds to the last row of the augmented matrix is

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

This equation can be omitted since it imposes no restrictions on x, y, and z; hence, the linear system corresponding to the augmented matrix is

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

In general, the variables in a linear system that correspond to the leading l's in its augmented matrix are called the ***leading variables***, and the remaining variables are called the ***free variables***. In this case the leading variables are x and y, and the variable z is the only free variable. Solving for the leading variables in terms of the free variables gives

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

From these equations we see that the free variable z can be treated as a parameter and assigned an arbitrary value t, which then determines values for x and y. Thus, the solution set can be represented by the parametric equations

$$x = -1 - 3t, \quad y = 2 + 4t, \quad z = t$$

By substituting various values for t in these equations we can obtain various solutions of the system. For example, setting $t = 0$ yields the solution

$$x = -1, \quad y = 2, \quad z = 0$$

and setting $t = 1$ yields the solution

$$x = -4, \quad y = 6, \quad z = 1$$

Solution (*c*) As explained in part (b), we can omit the equations corresponding to the zero rows, in which case the linear system associated with the augmented matrix consists of the single equation

$$x - 5y + z = 4 \tag{1}$$

from which we see that the solution set is a plane in three-dimensional space. Although (1) is a valid form of the solution set, there are many applications in which it is preferable to express the solution set in parametric form. We can convert (1) to parametric form by solving for the leading variable x in terms of the free variables y and z to obtain

$$x = 4 + 5y - z$$

From this equation we see that the free variables can be assigned arbitrary values, say $y = s$ and $z = t$, which then determine the value of x. Thus, the solution set can be expressed parametrically as

$$x = 4 + 5s - t, \quad y = s, \quad z = t \tag{2}$$

We will usually denote parameters in a general solution by the letters $r, s, t, \ldots$, but any letters that do not conflict with the names of the unknowns can be used. For systems with more than three unknowns, subscripted letters such as $t_1, t_2, t_3, \ldots$ are convenient.

Formulas, such as (2), that express the solution set of a linear system parametrically have some associated terminology.

Definition 1

If a linear system has infinitely many solutions, then a set of parametric equations from which all solutions can be obtained by assigning numerical values to the parameters is called a ***general solution*** of the system.

Thus, for example, Formula (2) is a general solution of system (iii) in the previous example.

Elimination Methods

We have just seen how easy it is to solve a system of linear equations once its augmented matrix is in reduced row echelon form. Now we will give a step-by-step *algorithm* that can be used to reduce any matrix to reduced row echelon form. As we state each step in the algorithm, we will illustrate the idea by reducing the following matrix to reduced row echelon form.

$$\begin{bmatrix} 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -10 & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix}$$

Step 1. Locate the leftmost column that does not consist entirely of zeros.

$$\begin{bmatrix} 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -10 & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix}$$

Leftmost nonzero column

Step 2. Interchange the top row with another row, if necessary, to bring a nonzero entry to the top of the column found in Step 1.

$$\begin{bmatrix} 2 & 4 & -10 & 6 & 12 & 28 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix}$$

The first and second rows in the preceding matrix were interchanged.

Step 3. If the entry that is now at the top of the column found in Step 1 is a, multiply the first row by $1/a$ in order to introduce a leading 1.

$$\begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix}$$

The first row of the preceding matrix was multiplied by $\frac{1}{2}$.

Step 4. Add suitable multiples of the top row to the rows below so that all entries below the leading 1 become zeros.

$$\begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 0 & 0 & 5 & 0 & -17 & -29 \end{bmatrix}$$

-2 times the first row of the preceding matrix was added to the third row.

Step 5. Now cover the top row in the matrix and begin again with Step 1 applied to the submatrix that remains. Continue in this way until the *entire* matrix is in row echelon form.

$$\begin{array}{c}\left[\begin{array}{rrrrrr} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 0 & 0 & 5 & 0 & -17 & -29 \end{array}\right] \\ \uparrow \text{ Leftmost nonzero column in the submatrix} \end{array}$$

$$\left[\begin{array}{rrrrrr} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\ 0 & 0 & 5 & 0 & -17 & -29 \end{array}\right] \longleftarrow$$

The first row in the submatrix was multiplied by $-\frac{1}{2}$ to introduce a leading 1.

$$\left[\begin{array}{rrrrrr} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 1 \end{array}\right] \longleftarrow$$

-5 times the first row of the submatrix was added to the second row of the submatrix to introduce a zero below the leading 1.

$$\begin{array}{c}\left[\begin{array}{rrrrrr} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 1 \end{array}\right] \longleftarrow \\ \uparrow \text{ Leftmost nonzero column in the new submatrix} \end{array}$$

The top row in the submatrix was covered, and we returned again to Step 1.

$$\left[\begin{array}{rrrrrr} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right] \longleftarrow$$

The first (and only) row in the new submatrix was multiplied by 2 to introduce a leading 1.

The *entire* matrix is now in row echelon form. To find the reduced row echelon form we need the following additional step.

Step 6. Beginning with the last nonzero row and working upward, add suitable multiples of each row to the rows above to introduce zeros above the leading 1's.

$$\left[\begin{array}{rrrrrr} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right] \longleftarrow$$

$\frac{7}{2}$ times the third row of the preceding matrix was added to the second row.

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

-6 times the third row was added to the first row.

$$\left[\begin{array}{rrrrrr} 1 & 2 & 0 & 3 & 0 & 7 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right] \longleftarrow$$

5 times the second row was added to the first row.

The last matrix is in reduced row echelon form.

The algorithm we have just described for reducing a matrix to reduced row echelon form is called ***Gauss–Jordan elimination***. It consists of two parts, a ***forward phase*** in which zeros are introduced below the leading 1's and a ***backward phase*** in which zeros are introduced above the leading 1's. If only the forward phase is used, then the procedure produces a row echelon form and is called ***Gaussian elimination***. For example, in the preceding computations a row echelon form was obtained at the end of Step 5.

Historical Note

Carl Friedrich Gauss (1777–1855)

Wilhelm Jordan (1842–1899)

Although versions of Gaussian elimination were known much earlier, its importance in scientific computation became clear when the great German mathematician Carl Friedrich Gauss used it to help compute the orbit of the asteroid Ceres from limited data. What happened was this: On January 1, 1801 the Sicilian astronomer and Catholic priest Giuseppe Piazzi (1746–1826) noticed a dim celestial object that he believed might be a "missing planet." He named the object Ceres and made a limited number of positional observations but then lost the object as it neared the Sun. Gauss, then only 24 years old, undertook the problem of computing the orbit of Ceres from the limited data using a technique called "least squares," the equations of which he solved by the method that we now call "Gaussian elimination." The work of Gauss created a sensation when Ceres reappeared a year later in the constellation Virgo at almost the precise position that he predicted! The basic idea of the method was further popularized by the German engineer Wilhelm Jordan in his book on geodesy (the science of measuring Earth shapes) entitled *Handbuch der Vermessungskunde* and published in 1888.

[Images: Photo Inc/Photo Researchers/Getty Images (Gauss); https://en.wikipedia.org/wiki/Andrey_Markov#/media/File:Andrei_Markov.jpg. Public domain. (Jordan)]

EXAMPLE 5 | Gauss–Jordan Elimination

Solve by Gauss–Jordan elimination.

$$\begin{aligned}
x_1 + 3x_2 - 2x_3 \phantom{{}-2x_4} + 2x_5 \phantom{{}-3x_6} &= 0\\
2x_1 + 6x_2 - 5x_3 - 2x_4 + 4x_5 - 3x_6 &= -1\\
5x_3 + 10x_4 \phantom{{}+4x_5} + 15x_6 &= 5\\
2x_1 + 6x_2 \phantom{{}-5x_3} + 8x_4 + 4x_5 + 18x_6 &= 6
\end{aligned}$$

Solution The augmented matrix for the system is

$$\begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0\\ 2 & 6 & -5 & -2 & 4 & -3 & -1\\ 0 & 0 & 5 & 10 & 0 & 15 & 5\\ 2 & 6 & 0 & 8 & 4 & 18 & 6 \end{bmatrix}$$

Adding -2 times the first row to the second and fourth rows gives

$$\begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0\\ 0 & 0 & -1 & -2 & 0 & -3 & -1\\ 0 & 0 & 5 & 10 & 0 & 15 & 5\\ 0 & 0 & 4 & 8 & 0 & 18 & 6 \end{bmatrix}$$

Multiplying the second row by -1 and then adding -5 times the new second row to the third row and -4 times the new second row to the fourth row gives

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

Interchanging the third and fourth rows and then multiplying the third row of the resulting matrix by $\frac{1}{6}$ gives the row echelon form

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

This completes the forward phase since there are zeros below the leading 1's.

Adding -3 times the third row to the second row and then adding 2 times the second row of the resulting matrix to the first row yields the reduced row echelon form

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

This completes the backward phase since there are zeros above the leading 1's.

The corresponding system of equations is

$$\begin{aligned} x_1 + 3x_2 \quad\quad + 4x_4 + 2x_5 \quad\quad &= 0 \\ x_3 + 2x_4 \quad\quad\quad\quad &= 0 \\ x_6 &= \tfrac{1}{3} \end{aligned} \tag{3}$$

Note that in constructing the linear system in (3) we ignored the row of zeros in the corresponding augmented matrix. Why is this justified?

Solving for the leading variables, we obtain

$$\begin{aligned} x_1 &= -3x_2 - 4x_4 - 2x_5 \\ x_3 &= -2x_4 \\ x_6 &= \tfrac{1}{3} \end{aligned}$$

Finally, we express the general solution of the system parametrically by assigning the free variables x_2, x_4, and x_5 arbitrary values r, s, and t, respectively. This yields

$$x_1 = -3r - 4s - 2t, \quad x_2 = r, \quad x_3 = -2s, \quad x_4 = s, \quad x_5 = t, \quad x_6 = \tfrac{1}{3}$$

Homogeneous Linear Systems

A system of linear equations is said to be ***homogeneous*** if the constant terms are all zero; that is, the system has the form

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

Every homogeneous system of linear equations is consistent because all such systems have $x_1 = 0, x_2 = 0, \ldots, x_n = 0$ as a solution. This solution is called the ***trivial solution***; if there are other solutions, they are called ***nontrivial solutions***.

Because a homogeneous linear system always has the trivial solution, there are only two possibilities for its solutions:

- The system has only the trivial solution.
- The system has infinitely many solutions in addition to the trivial solution.

In the special case of a homogeneous linear system of two equations in two unknowns, say

$$\begin{aligned} a_1x + b_1y &= 0 \quad [\boldsymbol{a_1, b_1} \text{ **not both zero**}] \\ a_2x + b_2y &= 0 \quad [\boldsymbol{a_2, b_2} \text{ **not both zero**}] \end{aligned}$$

the graphs of the equations are lines through the origin, and the trivial solution corresponds to the point of intersection at the origin (**Figure 1.2.1**).

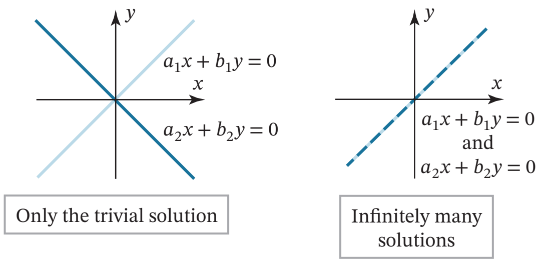


FIGURE 1.2.1

There is one case in which a homogeneous system is assured of having nontrivial solutions—namely, whenever the system involves more unknowns than equations. To see why, consider the following example of four equations in six unknowns.

EXAMPLE 6 | A Homogeneous System

Use Gauss–Jordan elimination to solve the homogeneous linear system

$$\begin{aligned} x_1 + 3x_2 - 2x_3 \qquad\quad + 2x_5 \qquad\quad &= 0 \\ 2x_1 + 6x_2 - 5x_3 - 2x_4 + 4x_5 - 3x_6 &= 0 \\ 5x_3 + 10x_4 \qquad\quad + 15x_6 &= 0 \\ 2x_1 + 6x_2 \qquad\quad + 8x_4 + 4x_5 + 18x_6 &= 0 \end{aligned} \tag{4}$$

Solution Observe that this system is the same as that in Example 5 except for the constants on the right side, which in this case are all zero. The augmented matrix for this system is

$$\begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 & 0 \\ 0 & 0 & 5 & 10 & 0 & 15 & 0 \\ 2 & 6 & 0 & 8 & 4 & 18 & 0 \end{bmatrix} \tag{5}$$

which is the same as that in Example 5 except for the entries in the last column, which are all zeros in this case. Thus, the reduced row echelon form of this matrix will be the same as that of the augmented matrix in Example 5, except for the last column. However, a moment's reflection will make it evident that a column of zeros is not changed by an elementary row operation, so the reduced row echelon form of (5) is

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

The corresponding system of equations is

$$\begin{aligned} x_1 + 3x_2 \qquad + 4x_4 + 2x_5 \qquad &= 0 \\ x_3 + 2x_4 \qquad\qquad &= 0 \\ x_6 &= 0 \end{aligned}$$

Solving for the leading variables, we obtain

$$\begin{aligned} x_1 &= -3x_2 - 4x_4 - 2x_5 \\ x_3 &= -2x_4 \\ x_6 &= 0 \end{aligned} \tag{7}$$

If we now assign the free variables x_2, x_4, and x_5 arbitrary values r, s, and t, respectively, then we can express the solution set parametrically as

$$x_1 = -3r - 4s - 2t, \quad x_2 = r, \quad x_3 = -2s, \quad x_4 = s, \quad x_5 = t, \quad x_6 = 0$$

Note that the trivial solution results when $r = s = t = 0$.

Free Variables in Homogeneous Linear Systems

Example 6 illustrates two important points about solving homogeneous linear systems:

1. Elementary row operations do not alter columns of zeros in a matrix, so the reduced row echelon form of the augmented matrix for a homogeneous linear system has a final column of zeros. This implies that the linear system corresponding to the reduced row echelon form is homogeneous, just like the original system.
2. When we constructed the homogeneous linear system corresponding to augmented matrix (6), we ignored the row of zeros because the corresponding equation

$$0x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 + 0x_6 = 0$$

   does not impose any conditions on the unknowns. Thus, depending on whether or not the reduced row echelon form of the augmented matrix for a homogeneous linear system has any zero rows, the linear system corresponding to that reduced row echelon form will either have the same number of equations as the original system or it will have fewer.

Now consider a general homogeneous linear system with n unknowns, and suppose that the reduced row echelon form of the augmented matrix has r nonzero rows. Since each nonzero row has a leading 1, and since each leading 1 corresponds to a leading variable, the homogeneous system corresponding to the reduced row echelon form of the augmented matrix must have r leading variables and $n - r$ free variables. Thus, this system is of the form

$$\begin{array}{cccccc} x_{k_1} & & & + \sum(\ ) = 0 \\ & x_{k_2} & & + \sum(\ ) = 0 \\ & & \ddots & \vdots \\ & & x_{k_r} & + \sum(\ ) = 0 \end{array} \tag{8}$$

where in each equation the expression $\sum(\ )$ denotes a sum that involves the free variables, if any [see (7), for example]. In summary, we have the following result.

Theorem 1.2.1

Free Variable Theorem for Homogeneous Systems

If a homogeneous linear system has n unknowns, and if the reduced row echelon form of its augmented matrix has r nonzero rows, then the system has $n - r$ free variables.

Theorem 1.2.1 has an important implication for homogeneous linear systems with more unknowns than equations. Specifically, if a homogeneous linear system has m equations in n unknowns, and if $m < n$, then it must also be true that $r < n$ (why?). This being the case, the theorem implies that there is at least one free variable, and this implies that the system has infinitely many solutions. Thus, we have the following result.

Note that Theorem 1.2.2 applies only to homogeneous systems—a *nonhomogeneous* system with more unknowns than equations need not be consistent. However, we will prove later that if a nonhomogeneous system with more unknowns than equations is consistent, then it has infinitely many solutions.

Theorem 1.2.2

A homogeneous linear system with more unknowns than equations has infinitely many solutions.

In retrospect, we could have anticipated that the homogeneous system in Example 6 would have infinitely many solutions since it has four equations in six unknowns.

Gaussian Elimination and Back-Substitution

For small linear systems that are solved by hand (such as most of those in this text), Gauss–Jordan elimination (reduction to reduced row echelon form) is a good procedure to use. However, for large linear systems that require a computer solution, it is generally more efficient to use Gaussian elimination (reduction to row echelon form) followed by a technique known as ***back-substitution*** to complete the process of solving the system. The next example illustrates this technique.

EXAMPLE 7 | Example 5 Solved by Back-Substitution

From the computations in Example 5, a row echelon form of the augmented matrix is

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

To solve the corresponding system of equations

$$\begin{aligned} x_1 + 3x_2 - 2x_3 + 2x_5 &= 0 \\ x_3 + 2x_4 + 3x_6 &= 1 \\ x_6 &= \tfrac{1}{3} \end{aligned}$$

we proceed as follows:

Step 1. Solve the equations for the leading variables.

$$\begin{aligned} x_1 &= -3x_2 + 2x_3 - 2x_5 \\ x_3 &= 1 - 2x_4 - 3x_6 \\ x_6 &= \tfrac{1}{3} \end{aligned}$$

Step 2. Beginning with the bottom equation and working upward, successively substitute each equation into all the equations above it.

Substituting $x_6 = \frac{1}{3}$ into the second equation yields

$$\begin{aligned} x_1 &= -3x_2 + 2x_3 - 2x_5 \\ x_3 &= -2x_4 \\ x_6 &= \tfrac{1}{3} \end{aligned}$$

Substituting $x_3 = -2x_4$ into the first equation yields

$$\begin{aligned} x_1 &= -3x_2 - 4x_4 - 2x_5 \\ x_3 &= -2x_4 \\ x_6 &= \tfrac{1}{3} \end{aligned}$$

Step 3. Assign arbitrary values to the free variables, if any.

If we now assign x_2, x_4, and x_5 the arbitrary values r, s, and t, respectively, the general solution is given by the formulas

$$x_1 = -3r - 4s - 2t, \quad x_2 = r, \quad x_3 = -2s, \quad x_4 = s, \quad x_5 = t, \quad x_6 = \tfrac{1}{3}$$

This agrees with the solution obtained in Example 5.

EXAMPLE 8 | Existence and Uniqueness of Solutions

Suppose that the matrices below are augmented matrices for linear systems in the unknowns x_1, x_2, x_3, and x_4. These matrices are all in row echelon form but not reduced row echelon form. Discuss the existence and uniqueness of solutions to the corresponding linear systems

$$(a)\ \begin{bmatrix} 1 & -3 & 7 & 2 & 5 \\ 0 & 1 & 2 & -4 & 1 \\ 0 & 0 & 1 & 6 & 9 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad (b)\ \begin{bmatrix} 1 & -3 & 7 & 2 & 5 \\ 0 & 1 & 2 & -4 & 1 \\ 0 & 0 & 1 & 6 & 9 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$(c)\ \begin{bmatrix} 1 & -3 & 7 & 2 & 5 \\ 0 & 1 & 2 & -4 & 1 \\ 0 & 0 & 1 & 6 & 9 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Solution (*a*) The last row corresponds to the equation

$$0x_1 + 0x_2 + 0x_3 + 0x_4 = 1$$

from which it is evident that the system is inconsistent.

Solution (*b*) The last row corresponds to the equation

$$0x_1 + 0x_2 + 0x_3 + 0x_4 = 0$$

which has no effect on the solution set. In the remaining three equations the variables x_1, x_2, and x_3 correspond to leading 1's and hence are leading variables. The variable x_4 is a free variable. With a little algebra, the leading variables can be expressed in terms of the free variable, and the free variable can be assigned an arbitrary value. Thus, the system must have infinitely many solutions.

Solution (*c*) The last row corresponds to the equation

$$x_4 = 0$$

which gives us a numerical value for x_4. If we substitute this value into the third equation, namely,

$$x_3 + 6x_4 = 9$$

we obtain $x_3 = 9$. You should now be able to see that if we continue this process and substitute the known values of x_3 and x_4 into the equation corresponding to the second row, we will obtain a unique numerical value for x_2; and if, finally, we substitute the known values of x_4, x_3, and x_2 into the equation corresponding to the first row, we will produce a unique numerical value for x_1. Thus, the system has a unique solution.

Some Facts About Echelon Forms

There are three facts about row echelon forms and reduced row echelon forms that are important to know but we will not prove:

1. Every matrix has a unique reduced row echelon form; that is, regardless of whether you use Gauss–Jordan elimination or some other sequence of elementary row operations, the same reduced row echelon form will result in the end.*
2. Row echelon forms are not unique; that is, different sequences of elementary row operations can result in different row echelon forms.

*A proof of this result can be found in the article "The Reduced Row Echelon Form of a Matrix Is Unique: A Simple Proof," by Thomas Yuster, *Mathematics Magazine*, Vol. 57, No. 2, 1984, pp. 93–94.

3. Although row echelon forms are not unique, the reduced row echelon form and all row echelon forms of a matrix A have the same number of zero rows, and the leading 1's always occur in the same positions. Those are called the ***pivot positions*** of A. The columns containing the leading 1's in a row echelon or reduced row echelon form of A are called the ***pivot columns*** of A, and the rows containing the leading 1's are called the ***pivot rows*** of A. A *nonzero* entry in a pivot position of A is called a ***pivot*** of A.

EXAMPLE 9 | Pivot Positions and Columns

Earlier in this section (immediately after Definition 1) we found a row echelon form of

$$A = \begin{bmatrix} 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -10 & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix} \quad \text{to be} \quad \begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}$$

The leading 1's occur in (row 1, column 1), (row 2, column 3), and (row 3, column 5). These are the pivot positions of A. The pivot columns of A are 1, 3, and 5, and the pivot rows are 1, 2, and 3. The pivots of A are the nonzero numbers in the pivot positions. These are marked by shaded rectangles in the following diagram.

$$A = \begin{bmatrix} 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & \boxed{-10} & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & \boxed{-5} & -1 \end{bmatrix}$$

Pivot columns

If A is the augmented matrix for a linear system, then the pivot columns identify the leading variables. As an illustration, in Example 5 the pivot columns are 1, 3, and 6, and the leading variables are x_1, x_3, and x_6.

Roundoff Error and Instability

There is often a gap between mathematical theory and its practical implementation—Gauss–Jordan elimination and Gaussian elimination being good examples. The problem is that computers generally approximate numbers, thereby introducing ***roundoff*** errors, so unless precautions are taken, successive calculations may degrade an answer to a degree that makes it useless. Algorithms in which this happens are called ***unstable***. There are various techniques for minimizing roundoff error and instability. For example, it can be shown that for large linear systems Gauss–Jordan elimination involves roughly 50% more operations than Gaussian elimination, so most computer algorithms are based on the latter method. Some of these matters will be considered in Chapter 9.

Exercise Set 1.2

In Exercises 1–2, determine whether the matrix is in row echelon form, reduced row echelon form, both, or neither.

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

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

f. $\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ **g.** $\begin{bmatrix} 1 & -7 & 5 & 5 \\ 0 & 1 & 3 & 2 \end{bmatrix}$

2. a. $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 0 \end{bmatrix}$ **c.** $\begin{bmatrix} 1 & 3 & 4 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

d. $\begin{bmatrix} 1 & 5 & -3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ **e.** $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

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

In Exercises 3–4, suppose that the augmented matrix for a linear system has been reduced by row operations to the given row echelon form. Identify the pivot rows and columns and solve the system.

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

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

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

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

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

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

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

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

In Exercises 5–8, solve the system by Gaussian elimination.

5. $\begin{aligned} x_1 + x_2 + 2x_3 &= 8 \\ -x_1 - 2x_2 + 3x_3 &= 1 \\ 3x_1 - 7x_2 + 4x_3 &= 10 \end{aligned}$

6. $\begin{aligned} 2x_1 + 2x_2 + 2x_3 &= 0 \\ -2x_1 + 5x_2 + 2x_3 &= 1 \\ 8x_1 + x_2 + 4x_3 &= -1 \end{aligned}$

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

8. $\begin{aligned} -2b + 3c &= 1 \\ 3a + 6b - 3c &= -2 \\ 6a + 6b + 3c &= 5 \end{aligned}$

In Exercises 9–12, solve the system by Gauss–Jordan elimination.

9. Exercise 5

10. Exercise 6

11. Exercise 7

12. Exercise 8

In Exercises 13–14, determine whether the homogeneous system has nontrivial solutions by inspection (without pencil and paper).

13. $\begin{aligned} 2x_1 - 3x_2 + 4x_3 - x_4 &= 0 \\ 7x_1 + x_2 - 8x_3 + 9x_4 &= 0 \\ 2x_1 + 8x_2 + x_3 - x_4 &= 0 \end{aligned}$

14. $\begin{aligned} x_1 + 3x_2 - x_3 &= 0 \\ x_2 - 8x_3 &= 0 \\ 4x_3 &= 0 \end{aligned}$

In Exercises 15–22, solve the given linear system by any method.

15. $\begin{aligned} 2x_1 + x_2 + 3x_3 &= 0 \\ x_1 + 2x_2 \qquad &= 0 \\ x_2 + x_3 &= 0 \end{aligned}$

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

17. $\begin{aligned} 3x_1 + x_2 + x_3 + x_4 &= 0 \\ 5x_1 - x_2 + x_3 - x_4 &= 0 \end{aligned}$

18. $\begin{aligned} v + 3w - 2x &= 0 \\ 2u + v - 4w + 3x &= 0 \\ 2u + 3v + 2w - x &= 0 \\ -4u - 3v + 5w - 4x &= 0 \end{aligned}$

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

20. $\begin{aligned} x_1 + 3x_2 \qquad + x_4 &= 0 \\ x_1 + 4x_2 + 2x_3 \qquad &= 0 \\ -2x_2 - 2x_3 - x_4 &= 0 \\ 2x_1 - 4x_2 + x_3 + x_4 &= 0 \\ x_1 - 2x_2 - x_3 + x_4 &= 0 \end{aligned}$

21. $\begin{aligned} 2I_1 - I_2 + 3I_3 + 4I_4 &= 9 \\ I_1 \qquad - 2I_3 + 7I_4 &= 11 \\ 3I_1 - 3I_2 + I_3 + 5I_4 &= 8 \\ 2I_1 + I_2 + 4I_3 + 4I_4 &= 10 \end{aligned}$

22. $\begin{aligned} Z_3 + Z_4 + Z_5 &= 0 \\ -Z_1 - Z_2 + 2Z_3 - 3Z_4 + Z_5 &= 0 \\ Z_1 + Z_2 - 2Z_3 \qquad - Z_5 &= 0 \\ 2Z_1 + 2Z_2 - Z_3 \qquad + Z_5 &= 0 \end{aligned}$

In each part of Exercises 23–24, the augmented matrix for a linear system is given in which the asterisk represents an unspecified real number. Determine whether the system is consistent, and if so whether the solution is unique. Answer "inconclusive" if there is not enough information to make a decision.

23. a. $\begin{bmatrix} 1 & * & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 1 & * \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & * & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 0 & 0 \end{bmatrix}$

c. $\begin{bmatrix} 1 & * & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 0 & 1 \end{bmatrix}$ **d.** $\begin{bmatrix} 1 & * & * & * \\ 0 & 0 & * & 0 \\ 0 & 0 & 1 & * \end{bmatrix}$

24. a. $\begin{bmatrix} 1 & * & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 1 & 1 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 0 & 0 & * \\ * & 1 & 0 & * \\ * & * & 1 & * \end{bmatrix}$

c. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & * & * & * \end{bmatrix}$ **d.** $\begin{bmatrix} 1 & * & * & * \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$

In Exercises 25–26, determine the values of a for which the system has no solutions, exactly one solution, or infinitely many solutions.

25. $\begin{aligned} x + 2y \qquad - 3z &= 4 \\ 3x - y + 5z &= 2 \\ 4x + y + (a^2 - 14)z &= a + 2 \end{aligned}$

26.
$$\begin{aligned} x + 2y + z &= 2 \\ 2x - 2y + 3z &= 1 \\ x + 2y - (a^2 - 3)z &= a \end{aligned}$$

In Exercises 27–28, what condition, if any, must a, b, and c satisfy for the linear system to be consistent?

27.
$$\begin{aligned} x + 3y - z &= a \\ x + y + 2z &= b \\ 2y - 3z &= c \end{aligned}$$

28.
$$\begin{aligned} x + 3y + z &= a \\ -x - 2y + z &= b \\ 3x + 7y - z &= c \end{aligned}$$

In Exercises 29–30, solve the following systems, where a, b, and c are constants.

29.
$$\begin{aligned} 2x + y &= a \\ 3x + 6y &= b \end{aligned}$$

30.
$$\begin{aligned} x_1 + x_2 + x_3 &= a \\ 2x_1 + 2x_3 &= b \\ 3x_2 + 3x_3 &= c \end{aligned}$$

31. Find two different row echelon forms of

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

This exercise shows that a matrix can have multiple row echelon forms.

32. Reduce

$$\begin{bmatrix} 2 & 1 & 3 \\ 0 & -2 & -29 \\ 3 & 4 & 5 \end{bmatrix}$$

to reduced row echelon form without introducing fractions at any intermediate stage.

33. Show that the following nonlinear system has 18 solutions if $0 \le \alpha \le 2\pi$, $0 \le \beta \le 2\pi$, and $0 \le \gamma \le 2\pi$.

$$\begin{aligned} \sin\alpha + 2\cos\beta + 3\tan\gamma &= 0 \\ 2\sin\alpha + 5\cos\beta + 3\tan\gamma &= 0 \\ -\sin\alpha - 5\cos\beta + 5\tan\gamma &= 0 \end{aligned}$$

[*Hint:* Begin by making the substitutions $x = \sin\alpha$, $y = \cos\beta$, and $z = \tan\gamma$.]

34. Solve the following system of nonlinear equations for the unknown angles α, β, and γ, where $0 \le \alpha \le 2\pi$, $0 \le \beta \le 2\pi$, and $0 \le \gamma < \pi$.

$$\begin{aligned} 2\sin\alpha - \cos\beta + 3\tan\gamma &= 3 \\ 4\sin\alpha + 2\cos\beta - 2\tan\gamma &= 2 \\ 6\sin\alpha - 3\cos\beta + \tan\gamma &= 9 \end{aligned}$$

35. Solve the following system of nonlinear equations for x, y, and z.

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

[*Hint:* Begin by making the substitutions $X = x^2$, $Y = y^2$, $Z = z^2$.]

36. Solve the following system for x, y, and z.

$$\begin{aligned} \frac{1}{x} + \frac{2}{y} - \frac{4}{z} &= 1 \\ \frac{2}{x} + \frac{3}{y} + \frac{8}{z} &= 0 \\ -\frac{1}{x} + \frac{9}{y} + \frac{10}{z} &= 5 \end{aligned}$$

37. Find the coefficients a, b, c, and d so that the curve shown in the accompanying figure is the graph of the equation $y = ax^3 + bx^2 + cx + d$.

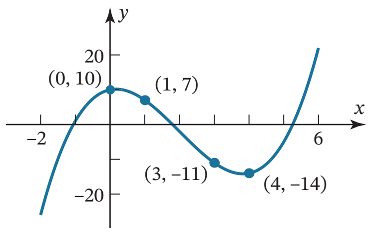


FIGURE Ex-37

38. Find the coefficients a, b, c, and d so that the circle shown in the accompanying figure is given by the equation $ax^2 + ay^2 + bx + cy + d = 0$.

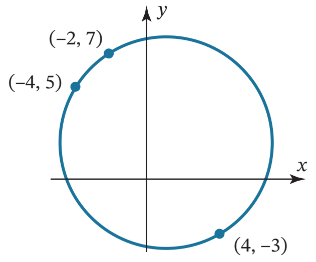


FIGURE Ex-38

39. If the linear system

$$\begin{aligned} a_1x + b_1y + c_1z &= 0 \\ a_2x - b_2y + c_2z &= 0 \\ a_3x + b_3y - c_3z &= 0 \end{aligned}$$

has only the trivial solution, what can be said about the solutions of the following system?

$$\begin{aligned} a_1x + b_1y + c_1z &= 3 \\ a_2x - b_2y + c_2z &= 7 \\ a_3x + b_3y - c_3z &= 11 \end{aligned}$$

40. **a.** If A is a matrix with three rows and five columns, then what is the maximum possible number of leading 1's in its reduced row echelon form?

b. If B is a matrix with three rows and six columns, then what is the maximum possible number of parameters in the general solution of the linear system with augmented matrix B?

c. If C is a matrix with five rows and three columns, then what is the minimum possible number of rows of zeros in any row echelon form of C?

41. Describe all possible reduced row echelon forms of

a. $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ **b.** $\begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & p & q \end{bmatrix}$

42. Consider the system of equations

$$\begin{aligned} ax + by &= 0 \\ cx + dy &= 0 \\ ex + fy &= 0 \end{aligned}$$

Discuss the relative positions of the lines $ax + by = 0$, $cx + dy = 0$, and $ex + fy = 0$ when the system has only the trivial solution and when it has nontrivial solutions.

Working with Proofs

43. **a.** Prove that if $ad - bc \neq 0$, then the reduced row echelon form of

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{is} \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

b. Use the result in part (a) to prove that if $ad - bc \neq 0$, then the linear system

$$\begin{aligned} ax + by &= k \\ cx + dy &= l \end{aligned}$$

has exactly one solution.

True-False Exercises

TF. In parts **(a)–(i)** determine whether the statement is true or false, and justify your answer.

a. If a matrix is in reduced row echelon form, then it is also in row echelon form.

b. If an elementary row operation is applied to a matrix that is in row echelon form, the resulting matrix will still be in row echelon form.

c. Every matrix has a unique row echelon form.

d. A homogeneous linear system in n unknowns whose corresponding augmented matrix has a reduced row echelon form with r leading 1's has $n - r$ free variables.

e. All leading 1's in a matrix in row echelon form must occur in different columns.

f. If every column of a matrix in row echelon form has a leading 1, then all entries that are not leading 1's are zero.

g. If a homogeneous linear system of n equations in n unknowns has a corresponding augmented matrix with a reduced row echelon form containing n leading 1's, then the linear system has only the trivial solution.

h. If the reduced row echelon form of the augmented matrix for a linear system has a row of zeros, then the system must have infinitely many solutions.

i. If a linear system has more unknowns than equations, then it must have infinitely many solutions.

Working with Technology

T1. Find the reduced row echelon form of the augmented matrix for the linear system

$$\begin{aligned} 6x_1 + x_2 \phantom{{}+3x_3} + 4x_4 &= -3 \\ -9x_1 + 2x_2 + 3x_3 - 8x_4 &= 1 \\ 7x_1 \phantom{{}+2x_2} - 4x_3 + 5x_4 &= 2 \end{aligned}$$

Use your result to determine whether the system is consistent and, if so, find its solution.

T2. Find values of the constants A, B, C, and D that make the following equation an identity (i.e., true for all values of x).

$$\frac{3x^3 + 4x^2 - 6x}{(x^2 + 2x + 2)(x^2 - 1)} = \frac{Ax + B}{x^2 + 2x + 2} + \frac{C}{x - 1} + \frac{D}{x + 1}$$

[*Hint:* Obtain a common denominator on the right, and then equate corresponding coefficients of the various powers of x in the two numerators. Students of calculus will recognize this as a problem in partial fractions.]

1.3 Matrices and Matrix Operations

Rectangular arrays of real numbers arise in contexts other than as augmented matrices for linear systems. In this section we will begin to study matrices as objects in their own right by defining operations of addition, subtraction, and multiplication on them.

Matrix Notation and Terminology

In Section 1.2 we used rectangular arrays of numbers, called *augmented matrices*, to abbreviate systems of linear equations. However, rectangular arrays of numbers occur in other contexts as well. For example, the following rectangular array with three rows and seven columns might describe the number of hours that a student spent studying three subjects during a certain week:

	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.
Math	2	3	2	4	1	4	2
History	0	3	1	4	3	2	2
Language	4	1	3	1	0	0	2

If we suppress the headings, then we are left with the following rectangular array of numbers with three rows and seven columns, called a "matrix":

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

More generally, we make the following definition.

Definition 1

A ***matrix*** is a rectangular array of numbers. The numbers in the array are called the ***entries*** of the matrix.

EXAMPLE 1 | Examples of Matrices

Some examples of matrices are

$$\begin{bmatrix} 1 & 2 \\ 3 & 0 \\ -1 & 4 \end{bmatrix}, \quad [2 \quad 1 \quad 0 \quad -3], \quad \begin{bmatrix} e & \pi & -\sqrt{2} \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad [4]$$

Matrix brackets are often omitted from 1×1 matrices, making it impossible to tell, for example, whether the symbol 4 denotes the number "four" or the matrix [4]. This rarely causes problems because it is usually possible to tell which is meant from the context.

The ***size*** of a matrix is described in terms of the number of rows (horizontal lines) and columns (vertical lines) it contains. For example, the first matrix in Example 1 has three rows and two columns, so its size is 3 by 2 (written 3×2). In a size description, the first number always denotes the number of rows, and the second denotes the number of columns. The remaining matrices in Example 1 have sizes 1×4, 3×3, 2×1, and 1×1, respectively.

A matrix with only one row, such as the second in Example 1, is called a ***row vector*** (or a ***row matrix***), and a matrix with only one column, such as the fourth in that example, is called a ***column vector*** (or a ***column matrix***). The fifth matrix in that example is both a row vector and a column vector.

We will use capital letters to denote matrices and lowercase letters to denote numerical quantities; thus we might write

$$A = \begin{bmatrix} 2 & 1 & 7 \\ 3 & 4 & 2 \end{bmatrix} \quad \text{or} \quad C = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

When discussing matrices, it is common to refer to numerical quantities as ***scalars***. Unless stated otherwise, *scalars will be real numbers*; complex scalars will be considered later in the text.

The entry that occurs in row i and column j of a matrix A will be denoted by a_{ij}. Thus a general 3×4 matrix might be written as

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

and a general $m \times n$ matrix as

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

When a compact notation is desired, matrix (1) can be written as

$$A = [a_{ij}]_{m \times n} \quad \text{or} \quad A = [a_{ij}]$$

the first notation being used when it is important in the discussion to know the size, and the second when the size need not be emphasized. Usually, we will match the letter denoting a matrix with the letter denoting its entries; thus, for a matrix B we would generally use b_{ij} for the entry in row i and column j, and for a matrix C we would use the notation c_{ij}.

The entry in row i and column j of a matrix A is also commonly denoted by the symbol $(A)_{ij}$. Thus, for matrix (1) above, we have

$$(A)_{ij} = a_{ij}$$

and for the matrix

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

we have $(A)_{11} = 2$, $(A)_{12} = -3$, $(A)_{21} = 7$, and $(A)_{22} = 0$.

Row and column vectors are of special importance, and it is common practice to denote them by boldface lowercase letters rather than capital letters. For such matrices, double subscripting of the entries is unnecessary. Thus a general $1 \times n$ row vector $\mathbf{a}$ and a general $m \times 1$ column vector $\mathbf{b}$ would be written as

$$\mathbf{a} = [a_1 \quad a_2 \quad \cdots \quad a_n] \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

A matrix A with n rows and n columns is called a ***square matrix of order n***, and the shaded entries $a_{11}, a_{22}, \ldots, a_{nn}$ in (2) are said to be on the ***main diagonal*** of A.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \tag{2}$$

Operations on Matrices

So far, we have used matrices to abbreviate the work in solving systems of linear equations. For other applications, however, it is desirable to develop an "arithmetic of matrices" in which matrices can be added, subtracted, and multiplied in a useful way. The remainder of this section will be devoted to developing this arithmetic.

Definition 2

Two matrices are defined to be ***equal*** if they have the same size and their corresponding entries are equal.

EXAMPLE 2 | Equality of Matrices

Consider the matrices

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

If $x = 5$, then $A = B$, but for all other values of x the matrices A and B are not equal, since not all of their corresponding entries are the same. There is no value of x for which $A = C$ since A and C have different sizes.

Definition 3

If A and B are matrices of the same size, then the ***sum*** $A + B$ is the matrix obtained by adding the entries of B to the corresponding entries of A, and the ***difference*** $A - B$ is the matrix obtained by subtracting the entries of B from the corresponding entries of A. Matrices of different sizes cannot be added or subtracted.

In matrix notation, if $A = [a_{ij}]$ and $B = [b_{ij}]$ have the same size, then

$$\boxed{(A + B)_{ij} = (A)_{ij} + (B)_{ij} = a_{ij} + b_{ij} \quad \text{and} \quad (A - B)_{ij} = (A)_{ij} - (B)_{ij} = a_{ij} - b_{ij}}$$

The equality of two matrices

$$A = [a_{ij}] \quad \text{and} \quad B = [b_{ij}]$$

of the same size can be expressed either by writing

$$(A)_{ij} = (B)_{ij}$$

or by writing

$$a_{ij} = b_{ij}$$

EXAMPLE 3 | Addition and Subtraction

Consider the matrices

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

Then

$$A + B = \begin{bmatrix} -2 & 4 & 5 & 4 \\ 1 & 2 & 2 & 3 \\ 7 & 0 & 3 & 5 \end{bmatrix} \quad \text{and} \quad A - B = \begin{bmatrix} 6 & -2 & -5 & 2 \\ -3 & -2 & 2 & 5 \\ 1 & -4 & 11 & -5 \end{bmatrix}$$

The expressions $A + C$, $B + C$, $A - C$, and $B - C$ are undefined.

Definition 4

If A is any matrix and c is any scalar, then the ***product*** cA is the matrix obtained by multiplying each entry of the matrix A by c. The matrix cA is said to be a ***scalar multiple*** of A.

In matrix notation, if $A = [a_{ij}]$, then

$$\boxed{(cA)_{ij} = c(A)_{ij} = ca_{ij}}$$

EXAMPLE 4 | Scalar Multiples

For the matrices

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 3 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 2 & 7 \\ -1 & 3 & -5 \end{bmatrix}, \quad C = \begin{bmatrix} 9 & -6 & 3 \\ 3 & 0 & 12 \end{bmatrix}$$

we have

$$2A = \begin{bmatrix} 4 & 6 & 8 \\ 2 & 6 & 2 \end{bmatrix}, \quad (-1)B = \begin{bmatrix} 0 & -2 & -7 \\ 1 & -3 & 5 \end{bmatrix}, \quad \tfrac{1}{3}C = \begin{bmatrix} 3 & -2 & 1 \\ 1 & 0 & 4 \end{bmatrix}$$

It is common practice to denote $(-1)B$ by $-B$.

Thus far we have defined multiplication of a matrix by a scalar but not the multiplication of two matrices. Since matrices are added by adding corresponding entries and subtracted by subtracting corresponding entries, it would seem natural to define multiplication of matrices by multiplying corresponding entries. However, it turns out that such a definition would not be very useful. Experience has led mathematicians to the following definition, the motivation for which will be given later in this chapter.

Definition 5

If A is an $m \times r$ matrix and B is an $r \times n$ matrix, then the ***product*** AB is the $m \times n$ matrix whose entries are determined as follows: To find the entry in row i and column j of AB, single out row i from the matrix A and column j from the matrix B. Multiply the corresponding entries from the row and column together, and then add the resulting products.

EXAMPLE 5 | Multiplying Matrices

Consider the matrices

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

Since A is a 2×3 matrix and B is a 3×4 matrix, the product AB is a 2×4 matrix. To determine, for example, the entry in row 2 and column 3 of AB, we single out row 2 from A and column 3 from B. Then, as illustrated below, we multiply corresponding entries together and add up these products.

$$\begin{bmatrix} 1 & 2 & 4 \\ 2 & 6 & 0 \end{bmatrix} \begin{bmatrix} 4 & 1 & 4 & 3 \\ 0 & -1 & 3 & 1 \\ 2 & 7 & 5 & 2 \end{bmatrix} = \begin{bmatrix} \square & \square & \square & \square \\ \square & \square & \boxed{26} & \square \end{bmatrix}$$

$$\boxed{(2 \cdot 4) + (6 \cdot 3) + (0 \cdot 5) = 26}$$

The entry in row 1 and column 4 of AB is computed as follows:

$$\begin{bmatrix} 1 & 2 & 4 \\ 2 & 6 & 0 \end{bmatrix} \begin{bmatrix} 4 & 1 & 4 & 3 \\ 0 & -1 & 3 & 1 \\ 2 & 7 & 5 & 2 \end{bmatrix} = \begin{bmatrix} \square & \square & \square & \boxed{13} \\ \square & \square & \square & \square \end{bmatrix}$$

$$\boxed{(1 \cdot 3) + (2 \cdot 1) + (4 \cdot 2) = 13}$$

The computations for the remaining entries are

$$\begin{aligned}
(1\cdot 4) + (2\cdot 0) + (4\cdot 2) &= 12\\
(1\cdot 1) - (2\cdot 1) + (4\cdot 7) &= 27\\
(1\cdot 4) + (2\cdot 3) + (4\cdot 5) &= 30\\
(2\cdot 4) + (6\cdot 0) + (0\cdot 2) &= 8\\
(2\cdot 1) - (6\cdot 1) + (0\cdot 7) &= -4\\
(2\cdot 3) + (6\cdot 1) + (0\cdot 2) &= 12
\end{aligned}
\qquad
AB = \begin{bmatrix} 12 & 27 & 30 & 13\\ 8 & -4 & 26 & 12 \end{bmatrix}$$

The definition of matrix multiplication requires that the number of columns of the first factor A be the same as the number of rows of the second factor B in order to form the product AB. If this condition is not satisfied, the product is undefined. A convenient way to determine whether a product of two matrices is defined is to write down the size of the first factor and, to the right of it, write down the size of the second factor. If, as in (3), the inside numbers are the same, then the product is defined. The outside numbers then give the size of the product.

$$\begin{array}{ccccc} A & & B & & AB \\ m \times \underbrace{r} & & \underbrace{r} \times n & = & m \times n \end{array} \tag{3}$$

Inside

Outside

EXAMPLE 6 | Determining Whether a Product Is Defined

Suppose that A, B, and C are matrices with the following sizes:

$$\begin{array}{ccc} A & B & C \\ 3\times 4 & 4\times 7 & 7\times 3 \end{array}$$

Then, AB is defined and is a 3×7 matrix; BC is defined and is a 4×3 matrix; and CA is defined and is a 7×4 matrix. The products AC, CB, and BA are all undefined.

In general, if $A = [a_{ij}]$ is an $m\times r$ matrix and $B = [b_{ij}]$ is an $r\times n$ matrix, then, as illustrated by the shading in the following display,

$$AB = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1r}\\ a_{21} & a_{22} & \cdots & a_{2r}\\ \vdots & \vdots & & \vdots\\ a_{i1} & a_{i2} & \cdots & a_{ir}\\ \vdots & \vdots & & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mr} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1n}\\ b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2n}\\ \vdots & \vdots & & \vdots & & \vdots\\ b_{r1} & b_{r2} & \cdots & b_{rj} & \cdots & b_{rn} \end{bmatrix} \tag{4}$$

the entry $(AB)_{ij}$ in row i and column j of AB is given by

$$\boxed{(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{ir}b_{rj}} \tag{5}$$

Formula (5) is called the ***row-column rule*** for matrix multiplication.

Partitioned Matrices

A matrix can be subdivided or ***partitioned*** into smaller matrices by inserting horizontal and vertical rules between selected rows and columns. For example, the following are three possible partitions of a general 3×4 matrix A—the first is a partition of A into four ***submatrices*** A_{11}, A_{12}, A_{21}, and A_{22}; the second is a partition of A into its row vectors $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$; and the third is a partition of A into its column vectors $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{c}_3$, and $\mathbf{c}_4$:

$$A = \left[\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

$$A = \left[\begin{array}{cccc} a_{11} & a_{12} & a_{13} & a_{14} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \end{array}\right] = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix}$$

$$A = \left[\begin{array}{c|c|c|c} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{array}\right] = [\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{c}_3 \quad \mathbf{c}_4]$$

Matrix Multiplication by Columns and by Rows

Partitioning has many uses, one of which is for finding particular rows or columns of a matrix product AB without computing the entire product. Specifically, the following formulas, whose proofs are left as exercises, show how individual column vectors of AB can be obtained by partitioning B into column vectors and how individual row vectors of AB can be obtained by partitioning A into row vectors.

$$\boxed{AB = A[\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_n] = [A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \cdots \quad A\mathbf{b}_n]} \tag{6}$$

(*AB* computed column by column)

$$\boxed{AB = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} B = \begin{bmatrix} \mathbf{a}_1 B \\ \mathbf{a}_2 B \\ \vdots \\ \mathbf{a}_m B \end{bmatrix}} \tag{7}$$

(*AB* computed row by row)

In words, these formulas state that

$$\boxed{j\text{th column vector of } AB = A[j\text{th column vector of } B]} \tag{8}$$

$$\boxed{i\text{th row vector of } AB = [i\text{th row vector of } A]B} \tag{9}$$

Historical Note

Gotthold Eisenstein (1823–1852)

The concept of matrix multiplication is due to the German mathematician Gotthold Eisenstein, who introduced the idea around 1844 to simplify the process of making substitutions in linear systems. The idea was then expanded on and formalized by Arthur Cayley (see p. 36) in his *Memoir on the Theory of Matrices* that was published in 1858. Eisenstein was a pupil of Gauss, who ranked him as the equal of Isaac Newton and Archimedes. However, Eisenstein, suffering from bad health his entire life, died at age 30, so his potential was never realized.

[Image: University of St Andrews/Wikipedia]

EXAMPLE 7 | Example 5 Revisited

If A and B are the matrices in Example 5, then from (8) the second column vector of AB can be obtained by the computation

$$\begin{bmatrix} 1 & 2 & 4 \\ 2 & 6 & 0 \end{bmatrix} \underset{\substack{\text{Second column} \\ \text{of } B}}{\begin{bmatrix} 1 \\ -1 \\ 7 \end{bmatrix}} = \underset{\substack{\text{Second column} \\ \text{of } AB}}{\begin{bmatrix} 27 \\ -4 \end{bmatrix}}$$

and from (9) the first row vector of AB can be obtained by the computation

$$\underset{\text{First row of } A}{\begin{bmatrix} 1 & 2 & 4 \end{bmatrix}} \begin{bmatrix} 4 & 1 & 4 & 3 \\ 0 & -1 & 3 & 1 \\ 2 & 7 & 5 & 2 \end{bmatrix} = \underset{\text{First row of } AB}{\begin{bmatrix} 12 & 27 & 30 & 13 \end{bmatrix}}$$

Matrix Products as Linear Combinations

The following definition provides yet another way of thinking about matrix multiplication.

Definition 6

If $A_1, A_2, \ldots, A_r$ are matrices of the same size, and if $c_1, c_2, \ldots, c_r$ are scalars, then an expression of the form

$$c_1A_1 + c_2A_2 + \cdots + c_rA_r$$

is called a ***linear combination*** of $A_1, A_2, \ldots, A_r$ with ***coefficients*** $c_1, c_2, \ldots, c_r$.

To see how matrix products can be viewed as linear combinations, let A be an $m \times n$ matrix and $\mathbf{x}$ an $n \times 1$ column vector, say

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Then

$$A\mathbf{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} = x_1\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2\begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n\begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} \tag{10}$$

This proves the following theorem.

Theorem 1.3.1

If A is an $m \times n$ matrix, and if $\mathbf{x}$ is an $n \times 1$ column vector, then the product $A\mathbf{x}$ can be expressed as a linear combination of the column vectors of A in which the coefficients are the entries of $\mathbf{x}$.

EXAMPLE 8 | Matrix Products as Linear Combinations

The matrix product

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

can be written as the following linear combination of column vectors:

$$2\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} - 1\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} + 3\begin{bmatrix} 2 \\ -3 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ -9 \\ -3 \end{bmatrix}$$

EXAMPLE 9 | Columns of a Product AB as Linear Combinations

We showed in Example 5 that

$$AB = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 6 & 0 \end{bmatrix} \begin{bmatrix} 4 & 1 & 4 & 3 \\ 0 & -1 & 3 & 1 \\ 2 & 7 & 5 & 2 \end{bmatrix} = \begin{bmatrix} 12 & 27 & 30 & 13 \\ 8 & -4 & 26 & 12 \end{bmatrix}$$

It follows from Formula (6) and Theorem 1.3.1 that the jth column vector of AB can be expressed as a linear combination of the column vectors of A in which the coefficients in the linear combination are the entries from the jth column of B. The computations are as follows:

$$\begin{bmatrix} 12 \\ 8 \end{bmatrix} = 4\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0\begin{bmatrix} 2 \\ 6 \end{bmatrix} + 2\begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 27 \\ -4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 6 \end{bmatrix} + 7\begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 30 \\ 26 \end{bmatrix} = 4\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 3\begin{bmatrix} 2 \\ 6 \end{bmatrix} + 5\begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 13 \\ 12 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 6 \end{bmatrix} + 2\begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

Column-Row Expansion

Partitioning provides yet another way to view matrix multiplication. Specifically, suppose that an $m \times r$ matrix A is partitioned into its r column vectors $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_r$ (each of size $m \times 1$) and an $r \times n$ matrix B is partitioned into its r row vectors $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_r$ (each of size $1 \times n$). Each term in the sum

$$\mathbf{c}_1\mathbf{r}_1 + \mathbf{c}_2\mathbf{r}_2 + \cdots + \mathbf{c}_r\mathbf{r}_r$$

has size $m \times n$ so the sum itself is an $m \times n$ matrix. We leave it as an exercise for you to verify that the entry in row i and column j of the sum is given by the expression on the right side of Formula (5), from which it follows that

$$\boxed{AB = \mathbf{c}_1\mathbf{r}_1 + \mathbf{c}_2\mathbf{r}_2 + \cdots + \mathbf{c}_r\mathbf{r}_r} \tag{11}$$

We call (11) the ***column-row expansion*** of AB.

EXAMPLE 10 | Column-Row Expansion

Find the column-row expansion of the product

$$AB = \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} 2 & 0 & 4 \\ -3 & 5 & 1 \end{bmatrix} \tag{12}$$

Solution The column vectors of A and the row vectors of B are, respectively,

$$\mathbf{c}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}; \quad \mathbf{r}_1 = \begin{bmatrix} 2 & 0 & 4 \end{bmatrix}, \quad \mathbf{r}_2 = \begin{bmatrix} -3 & 5 & 1 \end{bmatrix}$$

Thus, it follows from (11) that the column-row expansion of AB is

$$\begin{aligned} AB &= \begin{bmatrix} 1 \\ 2 \end{bmatrix}\begin{bmatrix} 2 & 0 & 4 \end{bmatrix} + \begin{bmatrix} 3 \\ -1 \end{bmatrix}\begin{bmatrix} -3 & 5 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 & 4 \\ 4 & 0 & 8 \end{bmatrix} + \begin{bmatrix} -9 & 15 & 3 \\ 3 & -5 & -1 \end{bmatrix} \end{aligned} \tag{13}$$

As a check, we leave it for you to confirm that the product in (12) and the sum in (13) both yield

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

Summarizing Matrix Multiplication

Putting it all together, we have given five different ways to compute a matrix product, each of which has its own use:

1. Entry by entry (Definition 5)
2. Row-column method (Formula (5))
3. Column by column (Formula (6))
4. Row by row (Formula (7))
5. Column-row expansion (Formula (11))

Matrix Form of a Linear System

Matrix multiplication has an important application to systems of linear equations. Consider a system of m linear equations in n unknowns:

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

Since two matrices are equal if and only if their corresponding entries are equal, we can replace the m equations in this system by the single matrix equation

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

The $m \times 1$ matrix on the left side of this equation can be written as a product to give

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

If we designate these matrices by A, $\mathbf{x}$, and $\mathbf{b}$, respectively, then we can replace the original system of m equations in n unknowns by the single matrix equation

$$\boxed{A\mathbf{x} = \mathbf{b}}$$

The matrix A in this equation is called the ***coefficient matrix*** of the system. The augmented matrix for the system is obtained by adjoining $\mathbf{b}$ to A as the last column; thus the augmented matrix is

$$[A \mid \mathbf{b}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

The vertical partition line in the augmented matrix $[A \mid \mathbf{b}]$ is optional, but is a useful way of visually separating the coefficient matrix A from the column vector $\mathbf{b}$.

Transpose of a Matrix

We conclude this section by defining two matrix operations that have no analogs in the arithmetic of real numbers.

Definition 7

If A is any $m \times n$ matrix, then the ***transpose of A***, denoted by A^T, is defined to be the $n \times m$ matrix that results by interchanging the rows and columns of A; that is, the first column of A^T is the first row of A, the second column of A^T is the second row of A, and so forth.

EXAMPLE 11 | Some Transposes

The following are some examples of matrices and their transposes.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 3 \\ 1 & 4 \\ 5 & 6 \end{bmatrix}, \quad C = [1 \quad 3 \quad 5], \quad D = [4]$$

$$A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \\ a_{14} & a_{24} & a_{34} \end{bmatrix}, \quad B^T = \begin{bmatrix} 2 & 1 & 5 \\ 3 & 4 & 6 \end{bmatrix}, \quad C^T = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}, \quad D^T = [4]$$

Observe that not only are the columns of A^T the rows of A, but the rows of A^T are the columns of A. Thus the entry in row i and column j of A^T is the entry in row j and column i of A; that is,

$$\boxed{(A^T)_{ij} = (A)_{ji}} \tag{14}$$

Note the reversal of the subscripts.

In the special case where A is a square matrix, the transpose of A can be obtained by interchanging entries that are symmetrically positioned about the main diagonal. In (15) we see that A^T can also be obtained by "reflecting" A about its main diagonal.

$$A = \begin{bmatrix} 1 & -2 & 4 \\ 3 & 7 & 0 \\ -5 & 8 & 6 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -2 & 4 \\ 3 & 7 & 0 \\ -5 & 8 & 6 \end{bmatrix} \rightarrow A^T = \begin{bmatrix} 1 & 3 & -5 \\ -2 & 7 & 8 \\ 4 & 0 & 6 \end{bmatrix} \tag{15}$$

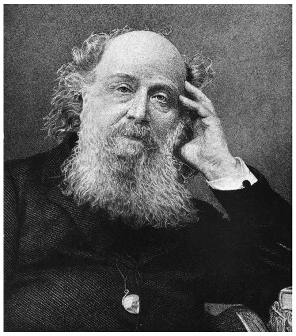
Interchange entries that are symmetrically positioned about the main diagonal.

Trace of a Matrix

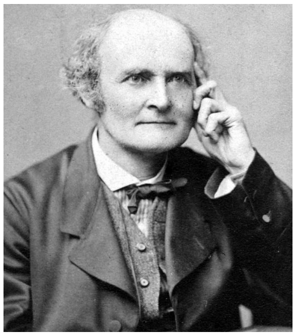
Definition 8

If A is a square matrix, then the ***trace of A***, denoted by $\operatorname{tr}(A)$, is defined to be the sum of the entries on the main diagonal of A. The trace of A is undefined if A is not a square matrix.

Historical Note



James Sylvester (1814–1897)



Arthur Cayley (1821–1895)

The term *matrix* was first used by the English mathematician James Sylvester, who defined the term in 1850 to be an "oblong arrangement of terms." Sylvester communicated his work on matrices to a fellow English mathematician and lawyer named Arthur Cayley, who then introduced some of the basic operations on matrices in a book entitled *Memoir on the Theory of Matrices* that was published in 1858. As a matter of interest, Sylvester, who was Jewish, did not get his college degree because he refused to sign a required oath to the Church of England. He was appointed to a chair at the University of Virginia in the United States but resigned after swatting a student with a stick because he was reading a newspaper in class. Sylvester, thinking he had killed the student, fled back to England on the first available ship. Fortunately, the student was not dead, just in shock!

[Images: © Bettmann/CORBIS (Sylvester); Wikipedia Commons (Cayley)]

EXAMPLE 12 | Trace

The following are examples of matrices and their traces.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 2 & 7 & 0 \\ 3 & 5 & -8 & 4 \\ 1 & 2 & 7 & -3 \\ 4 & -2 & 1 & 0 \end{bmatrix}$$

$$\boxed{\operatorname{tr}(A) = a_{11} + a_{22} + a_{33}} \qquad \boxed{\operatorname{tr}(B) = -1 + 5 + 7 + 0 = 11}$$

In the exercises you will have some practice working with the transpose and trace operations.

Exercise Set 1.3

In Exercises 1–2, suppose that A, B, C, D, and E are matrices with the following sizes:

A	B	C	D	E
(4×5)	(4×5)	(5×2)	(4×2)	(5×4)

In each part, determine whether the given matrix expression is defined. For those that are defined, give the size of the resulting matrix.

1. **a.** BA **b.** AB^T **c.** $AC + D$
d. $E(AC)$ **e.** $A - 3E^T$ **f.** $E(5B + A)$

2. **a.** CD^T **b.** DC **c.** $BC - 3D$
d. $D^T(BE)$ **e.** $B^TD + ED$ **f.** $BA^T + D$

In Exercises 3–6, use the following matrices to compute the indicated expression if it is defined.

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

$$D = \begin{bmatrix} 1 & 5 & 2 \\ -1 & 0 & 1 \\ 3 & 2 & 4 \end{bmatrix}, \quad E = \begin{bmatrix} 6 & 1 & 3 \\ -1 & 1 & 2 \\ 4 & 1 & 3 \end{bmatrix}$$

3. **a.** $D + E$ **b.** $D - E$ **c.** $5A$
d. $-7C$ **e.** $2B - C$ **f.** $4E - 2D$
g. $-3(D + 2E)$ **h.** $A - A$ **i.** $\operatorname{tr}(D)$
j. $\operatorname{tr}(D - 3E)$ **k.** $4\operatorname{tr}(7B)$ **l.** $\operatorname{tr}(A)$

4. **a.** $2A^T + C$ **b.** $D^T - E^T$ **c.** $(D - E)^T$
d. $B^T + 5C^T$ **e.** $\frac{1}{2}C^T - \frac{1}{4}A$ **f.** $B - B^T$
g. $2E^T - 3D^T$ **h.** $(2E^T - 3D^T)^T$ **i.** $(CD)E$
j. $C(BA)$ **k.** $\operatorname{tr}(DE^T)$ **l.** $\operatorname{tr}(BC)$

5. **a.** AB **b.** BA **c.** $(3E)D$
d. $(AB)C$ **e.** $A(BC)$ **f.** CC^T
g. $(DA)^T$ **h.** $(C^TB)A^T$ **i.** $\operatorname{tr}(DD^T)$
j. $\operatorname{tr}(4E^T - D)$ **k.** $\operatorname{tr}(C^TA^T + 2E^T)$ **l.** $\operatorname{tr}((EC^T)^TA)$

6. **a.** $(2D^T - E)A$ **b.** $(4B)C + 2B$
c. $(-AC)^T + 5D^T$ **d.** $(BA^T - 2C)^T$
e. $B^T(CC^T - A^TA)$ **f.** $D^TE^T - (ED)^T$

In Exercises 7–8, use the following matrices and either the row method or the column method, as appropriate, to find the indicated row or column.

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

7. **a.** the first row of AB **b.** the third row of AB
c. the second column of AB **d.** the first column of BA
e. the third row of AA **f.** the third column of AA

8. **a.** the first column of AB **b.** the third column of BB
c. the second row of BB **d.** the first column of AA
e. the third column of AB **f.** the first row of BA

In Exercises 9–10, use matrices A and B from Exercises 7–8.

9. **a.** Express each column vector of AA as a linear combination of the column vectors of A.

b. Express each column vector of BB as a linear combination of the column vectors of B.

10. **a.** Express each column vector of AB as a linear combination of the column vectors of A.

b. Express each column vector of BA as a linear combination of the column vectors of B.

In each part of Exercises 11–12, find matrices A, $\mathbf{x}$, and $\mathbf{b}$ that express the given linear system as a single matrix equation $A\mathbf{x} = \mathbf{b}$, and write out this matrix equation.

11. **a.**
$$\begin{aligned} 2x_1 - 3x_2 + 5x_3 &= 7 \\ 9x_1 - x_2 + x_3 &= -1 \\ x_1 + 5x_2 + 4x_3 &= 0 \end{aligned}$$

b.
$$\begin{aligned} 4x_1 \qquad - 3x_3 + x_4 &= 1 \\ 5x_1 + x_2 \qquad - 8x_4 &= 3 \\ 2x_1 - 5x_2 + 9x_3 - x_4 &= 0 \\ 3x_2 - x_3 + 7x_4 &= 2 \end{aligned}$$

12. **a.**
$$\begin{aligned} x_1 - 2x_2 + 3x_3 &= -3 \\ 2x_1 + x_2 \qquad &= 0 \\ -3x_2 + 4x_3 &= 1 \\ x_1 \qquad + x_3 &= 5 \end{aligned}$$

b.
$$\begin{aligned} 3x_1 + 3x_2 + 3x_3 &= -3 \\ -x_1 - 5x_2 - 2x_3 &= 3 \\ -4x_2 + x_3 &= 0 \end{aligned}$$

In each part of Exercises 13–14, express the matrix equation as a system of linear equations.

13. **a.** $\begin{bmatrix} 5 & 6 & -7 \\ -1 & -2 & 3 \\ 0 & 4 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix}$

b. $\begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 0 \\ 5 & -3 & -6 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ -9 \end{bmatrix}$

14. **a.** $\begin{bmatrix} 3 & -1 & 2 \\ 4 & 3 & 7 \\ -2 & 1 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}$

b. $\begin{bmatrix} 3 & -2 & 0 & 1 \\ 5 & 0 & 2 & -2 \\ 3 & 1 & 4 & 7 \\ -2 & 5 & 1 & 6 \end{bmatrix}\begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

In Exercises 15–16, find all values of k, if any, that satisfy the equation.

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

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

In Exercises 17–20, use the column-row expansion of AB to express this product as a sum of matrix products.

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

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

19. $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$

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

21. For the linear system in Example 5 of Section 1.2, express the general solution that we obtained in that example as a linear combination of column vectors that contain only numerical entries. [*Suggestion:* Rewrite the general solution as a single column vector, then write that column vector as a sum of column vectors each of which contains at most one parameter, and then factor out the parameters.]

22. Follow the directions of Exercise 21 for the linear system in Example 6 of Section 1.2.

In Exercises 23–24, solve the matrix equation for a, b, c, and d.

23. $\begin{bmatrix} a & 3 \\ -1 & a+b \end{bmatrix} = \begin{bmatrix} 4 & d-2c \\ d+2c & -2 \end{bmatrix}$

24. $\begin{bmatrix} a-b & b+a \\ 3d+c & 2d-c \end{bmatrix} = \begin{bmatrix} 8 & 1 \\ 7 & 6 \end{bmatrix}$

25. **a.** Show that if A has a row of zeros and B is any matrix for which AB is defined, then AB also has a row of zeros.

b. Find a similar result involving a column of zeros.

26. In each part, find a 6×6 matrix $[a_{ij}]$ that satisfies the stated condition. Make your answers as general as possible by using letters rather than specific numbers for the nonzero entries.

a. $a_{ij} = 0 \quad \text{if} \quad i \neq j$ **b.** $a_{ij} = 0 \quad \text{if} \quad i > j$

c. $a_{ij} = 0 \quad \text{if} \quad i < j$ **d.** $a_{ij} = 0 \quad \text{if} \quad |i - j| > 1$

In Exercises 27–28, how many 3×3 matrices A can you find for which the equation is satisfied for all choices of x, y, and z?

27. $A\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+y \\ x-y \\ 0 \end{bmatrix}$

28. $A\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} xy \\ 0 \\ 0 \end{bmatrix}$

29. A matrix B is said to be a ***square root*** of a matrix A if $BB = A$.

a. Find two square roots of $A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$.

b. How many different square roots can you find of $A = \begin{bmatrix} 5 & 0 \\ 0 & 9 \end{bmatrix}$?

c. Do you think that every 2×2 matrix has at least one square root? Explain your reasoning.

30. Let 0 denote a 2×2 matrix, each of whose entries is zero.

a. Is there a 2×2 matrix A such that $A \neq 0$ and $AA = 0$? Justify your answer.

b. Is there a 2×2 matrix A such that $A \neq 0$ and $AA = A$? Justify your answer.

31. Establish Formula (11) by using Formula (5) to show that

$$(AB)_{ij} = (\mathbf{c}_1\mathbf{r}_1 + \mathbf{c}_2\mathbf{r}_2 + \cdots + \mathbf{c}_r\mathbf{r}_r)_{ij}$$

32. Find a 4×4 matrix $A = [a_{ij}]$ whose entries satisfy the stated condition.

a. $a_{ij} = i + j$ **b.** $a_{ij} = i^{j-1}$

c. $a_{ij} = \begin{cases} 1 & \text{if} \quad |i-j| > 1 \\ -1 & \text{if} \quad |i-j| \leq 1 \end{cases}$

33. Suppose that type I items cost \$1 each, type II items cost \$2 each, and type III items cost \$3 each. Also, suppose that the accompanying table describes the number of items of each type purchased during the first four months of the year.

TABLE Ex-33

	Type I	Type II	Type III
Jan.	3	4	3
Feb.	5	6	0
Mar.	2	9	4
Apr.	1	1	7

What information is represented by the following product?

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

34. The accompanying table shows a record of May and June unit sales for a clothing store. Let M denote the 4×3 matrix of May sales and J the 4×3 matrix of June sales.

a. What does the matrix $M + J$ represent?

b. What does the matrix $M - J$ represent?

c. Find a column vector $\mathbf{x}$ for which $M\mathbf{x}$ provides a list of the number of shirts, jeans, suits, and raincoats sold in May.

d. Find a row vector $\mathbf{y}$ for which $\mathbf{y}M$ provides a list of the number of small, medium, and large items sold in May.

e. Using the matrices $\mathbf{x}$ and $\mathbf{y}$ that you found in parts (c) and (d), what does $\mathbf{y}M\mathbf{x}$ represent?

TABLE Ex-34

May Sales

	Small	Medium	Large
Shirts	45	60	75
Jeans	30	30	40
Suits	12	65	45
Raincoats	15	40	35

June Sales

	Small	Medium	Large
Shirts	30	33	40
Jeans	21	23	25
Suits	9	12	11
Raincoats	8	10	9

Working with Proofs

35. Prove: If A and B are $n \times n$ matrices, then

$$\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$$

36. a. Prove: If AB and BA are both defined, then AB and BA are square matrices.

b. Prove: If A is an $m \times n$ matrix and $A(BA)$ is defined, then B is an $n \times m$ matrix.

True-False Exercises

TF. In parts **(a)–(o)** determine whether the statement is true or false, and justify your answer.

a. The matrix $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ has no main diagonal.

b. An $m \times n$ matrix has m column vectors and n row vectors.

c. If A and B are 2×2 matrices, then $AB = BA$.

d. The ith row vector of a matrix product AB can be computed by multiplying A by the ith row vector of B.

e. For every matrix A, it is true that $(A^T)^T = A$.

f. If A and B are square matrices of the same order, then

$$\operatorname{tr}(AB) = \operatorname{tr}(A)\operatorname{tr}(B)$$

g. If A and B are square matrices of the same order, then

$$(AB)^T = A^T B^T$$

h. For every square matrix A, it is true that $\operatorname{tr}(A^T) = \operatorname{tr}(A)$.

i. If A is a 6×4 matrix and B is an $m \times n$ matrix such that $B^T A^T$ is a 2×6 matrix, then $m = 4$ and $n = 2$.

j. If A is an $n \times n$ matrix and c is a scalar, then $\operatorname{tr}(cA) = c\operatorname{tr}(A)$.

k. If A, B, and C are matrices of the same size such that $A - C = B - C$, then $A = B$.

l. If A, B, and C are square matrices of the same order such that $AC = BC$, then $A = B$.

m. If $AB + BA$ is defined, then A and B are square matrices of the same size.

n. If B has a column of zeros, then so does AB if this product is defined.

o. If B has a column of zeros, then so does BA if this product is defined.

Working with Technology

T1. a. Compute the product AB of the matrices in Example 5, and compare your answer to that in the text.

b. Use your technology utility to extract the columns of A and the rows of B, and then calculate the product AB by a column-row expansion.

T2. Suppose that a manufacturer uses Type I items at \$1.35 each, Type II items at \$2.15 each, and Type III items at \$3.95 each. Suppose also that the accompanying table describes the purchases of those items (in thousands of units) for the first quarter of the year. Find a matrix product, the computation of which produces a matrix that lists the manufacturer's expenditure in each month of the first quarter. Compute that product.

	Type I	Type II	Type III
Jan.	3.1	4.2	3.5
Feb.	5.1	6.8	0
Mar.	2.2	9.5	4.0
Apr.	1.0	1.0	7.4

1.4 Inverses; Algebraic Properties of Matrices

In this section we will discuss some of the algebraic properties of matrix operations. We will see that many of the basic rules of arithmetic for real numbers hold for matrices, but we will also see that some do not.

Properties of Matrix Addition and Scalar Multiplication

The following theorem lists the basic algebraic properties of the matrix operations.

Theorem 1.4.1

Properties of Matrix Arithmetic

Assuming that the sizes of the matrices are such that the indicated operations can be performed, the following rules of matrix arithmetic are valid.

(*a*) $A + B = B + A$ **[Commutative law for matrix addition]**

(*b*) $A + (B + C) = (A + B) + C$ **[Associative law for matrix addition]**

(*c*) $A(BC) = (AB)C$ **[Associative law for matrix multiplication]**

(*d*) $A(B + C) = AB + AC$ **[Left distributive law]**

(*e*) $(B + C)A = BA + CA$ **[Right distributive law]**

(*f*) $A(B - C) = AB - AC$

(*g*) $(B - C)A = BA - CA$

(*h*) $a(B + C) = aB + aC$

(*i*) $a(B - C) = aB - aC$

(*j*) $(a + b)C = aC + bC$

(*k*) $(a - b)C = aC - bC$

(*l*) $a(bC) = (ab)C$

(*m*) $a(BC) = (aB)C = B(aC)$

To prove any of the equalities in this theorem one must show that the matrix on the left side has the same size as that on the right and that the corresponding entries on the two sides are the same. Most of the proofs follow the same pattern, so we will prove part (*d*) as a sample. The proof of the associative law for multiplication is more complicated than the rest and is outlined in the exercises.

Proof (d) We must show that $A(B+C)$ and $AB+AC$ have the same size and that corresponding entries are equal. To form $A(B+C)$, the matrices B and C must have the same size, say $m \times n$, and the matrix A must then have m columns, so its size must be of the form $r \times m$. This makes $A(B+C)$ an $r \times n$ matrix. It follows that $AB+AC$ is also an $r \times n$ matrix and, consequently, $A(B+C)$ and $AB+AC$ have the same size.

Suppose that $A=[a_{ij}]$, $B=[b_{ij}]$, and $C=[c_{ij}]$. We want to show that corresponding entries of $A(B+C)$ and $AB+AC$ are equal; that is,

$$\big(A(B+C)\big)_{ij} = (AB+AC)_{ij}$$

for all values of i and j. But from the definitions of matrix addition and matrix multiplication, we have

$$\begin{aligned}\big(A(B+C)\big)_{ij} &= a_{i1}(b_{1j}+c_{1j}) + a_{i2}(b_{2j}+c_{2j}) + \cdots + a_{im}(b_{mj}+c_{mj}) \\ &= (a_{i1}b_{1j}+a_{i2}b_{2j}+\cdots+a_{im}b_{mj}) + (a_{i1}c_{1j}+a_{i2}c_{2j}+\cdots+a_{im}c_{mj}) \\ &= (AB)_{ij} + (AC)_{ij} = (AB+AC)_{ij} \quad \blacksquare\end{aligned}$$

There are three basic ways to prove that two matrices of the same size are equal—prove that corresponding entries are the same, prove that corresponding row vectors are the same, or prove that corresponding column vectors are the same.

Remark Although the operations of matrix addition and matrix multiplication were defined for pairs of matrices, associative laws (b) and (c) enable us to denote sums and products of three matrices as $A+B+C$ and ABC without inserting any parentheses. This is justified by the fact that no matter how parentheses are inserted, the associative laws guarantee that the same end result will be obtained. In general, *given any sum or any product of matrices, pairs of parentheses can be inserted or deleted anywhere within the expression without affecting the end result.*

EXAMPLE 1 | Associativity of Matrix Multiplication

As an illustration of the associative law for matrix multiplication, consider

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

Then

$$AB = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 5 \\ 20 & 13 \\ 2 & 1 \end{bmatrix} \quad \text{and} \quad BC = \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 10 & 9 \\ 4 & 3 \end{bmatrix}$$

Thus

$$(AB)C = \begin{bmatrix} 8 & 5 \\ 20 & 13 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 18 & 15 \\ 46 & 39 \\ 4 & 3 \end{bmatrix}$$

and

$$A(BC) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 10 & 9 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 18 & 15 \\ 46 & 39 \\ 4 & 3 \end{bmatrix}$$

so $(AB)C = A(BC)$, as guaranteed by Theorem 1.4.1(c).

Properties of Matrix Multiplication

Do not let Theorem 1.4.1 lull you into believing that *all* laws of real arithmetic carry over to matrix arithmetic. For example, you know that in real arithmetic it is always true that

$ab = ba$, which is called the *commutative law for multiplication*. In matrix arithmetic, however, the equality of AB and BA can fail for three possible reasons:

1. AB may be defined and BA may not (for example, if A is 2×3 and B is 3×4).
2. AB and BA may both be defined, but they may have different sizes (for example, if A is 2×3 and B is 3×2).
3. AB and BA may both be defined and have the same size, but the two products may be different (as illustrated in the next example).

EXAMPLE 2 | Order Matters in Matrix Multiplication

Consider the matrices

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

Multiplying gives

$$AB = \begin{bmatrix} -1 & -2 \\ 11 & 4 \end{bmatrix} \quad \text{and} \quad BA = \begin{bmatrix} 3 & 6 \\ -3 & 0 \end{bmatrix}$$

Thus, $AB \neq BA$.

Because, as this example shows, it is *not* generally true that $AB = BA$, we say that ***matrix multiplication is not commutative***. This does not preclude the possibility of equality in certain cases—it is just not true in general. In those special cases where there is equality we say that A and B ***commute***.

Zero Matrices

A matrix whose entries are all zero is called a ***zero matrix***. Some examples are

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

We will denote a zero matrix by 0 unless it is important to specify its size, in which case we will denote the $m \times n$ zero matrix by $0_{m \times n}$.

It should be evident that if A and 0 are matrices with the same size, then

$$A + 0 = 0 + A = A$$

Thus, 0 plays the same role in this matrix equation that the number 0 plays in the numerical equation $a + 0 = 0 + a = a$.

The following theorem lists the basic properties of zero matrices. Since the results should be self-evident, we will omit the formal proofs.

Theorem 1.4.2

Properties of Zero Matrices

If c is a scalar, and if the sizes of the matrices are such that the operations can be perfomed, then:

(*a*) $A + 0 = 0 + A = A$

(*b*) $A - 0 = A$

(*c*) $A - A = A + (-A) = 0$

(*d*) $0A = 0$

(*e*) If $cA = 0$, then $c = 0$ or $A = 0$.

Since we know that the commutative law of real arithmetic is not valid in matrix arithmetic, it should not be surprising that there are other rules that fail as well. For example, consider the following two laws of real arithmetic:

- If $ab = ac$ and $a \neq 0$, then $b = c$. **[The cancellation law]**
- If $ab = 0$, then at least one of the factors on the left is 0.

The next two examples show that these laws are not true in matrix arithmetic.

EXAMPLE 3 | Failure of the Cancellation Law

Consider the matrices

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

We leave it for you to confirm that

$$AB = AC = \begin{bmatrix} 3 & 4 \\ 6 & 8 \end{bmatrix}$$

Although $A \neq 0$, canceling A from both sides of the equation $AB = AC$ would lead to the incorrect conclusion that $B = C$. Thus, the cancellation law does not hold, in general, for matrix multiplication (though there may be particular cases where it is true).

EXAMPLE 4 | A Zero Product with Nonzero Factors

Here are two matrices for which $AB = 0$, but $A \neq 0$ and $B \neq 0$:

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

Identity Matrices

A square matrix with 1's on the main diagonal and zeros elsewhere is called an ***identity matrix***. Some examples are

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

An identity matrix is denoted by the letter I. If it is important to emphasize the size, we will write I_n for the $n \times n$ identity matrix.

To explain the role of identity matrices in matrix arithmetic, let us consider the effect of multiplying a general 2×3 matrix A on each side by an identity matrix. Multiplying on the right by the 3×3 identity matrix yields

$$AI_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = A$$

and multiplying on the left by the 2×2 identity matrix yields

$$I_2A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = A$$

The same result holds in general; that is, if A is any $m \times n$ matrix, then

$$\boxed{AI_n = A \quad \text{and} \quad I_mA = A}$$

Thus, the identity matrices play the same role in matrix arithmetic that the number 1 plays in the numerical equation $a \cdot 1 = 1 \cdot a = a$.

As the next theorem shows, identity matrices arise naturally as reduced row echelon forms of *square* matrices.

Theorem 1.4.3

If R is the reduced row echelon form of an $n \times n$ matrix A, then either R has at least one row of zeros or R is the identity matrix I_n.

Proof Suppose that the reduced row echelon form of A is

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

Either the last row in this matrix consists entirely of zeros or it does not. If not, the matrix contains no zero rows, and consequently each of the n rows has a leading entry of 1. Since these leading 1's occur progressively farther to the right as we move down the matrix, each of these 1's must occur on the main diagonal. Since the other entries in the same column as one of these 1's are zero, R must be I_n. Thus, either R has a row of zeros or $R = I_n$. ■

Inverse of a Matrix

In real arithmetic every nonzero number a has a reciprocal $a^{-1}(= 1/a)$ with the property

$$a \cdot a^{-1} = a^{-1} \cdot a = 1$$

The number a^{-1} is sometimes called the *multiplicative inverse* of a. Our next objective is to develop an analog of this result for matrix arithmetic. For this purpose we make the following definition.

Definition 1

If A is a square matrix, and if there exists a matrix B of the same size for which $AB = BA = I$, then A is said to be ***invertible*** (or ***nonsingular***) and B is called an ***inverse*** of A. If no such matrix B exists, then A is said to be ***singular***.

The relationship $AB = BA = I$ is not changed by interchanging A and B, so if A is invertible and B is an inverse of A, then it is also true that B is invertible, and A is an inverse of B. Thus, when $AB = BA = I$ we say that A and B are *inverses of one another*.

EXAMPLE 5 | An Invertible Matrix

Let

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

Then

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

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

Thus, A and B are invertible and each is an inverse of the other.

EXAMPLE 6 | A Class of Singular Matrices

A square matrix with a row or column of zeros is singular. To help understand why this is so, consider the matrix

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

To prove that A is singular we must show that there is no 3×3 matrix B such that

$$AB = BA = I$$

For this purpose let $\mathbf{c}_1, \mathbf{c}_2, \mathbf{0}$ be the column vectors of A. Thus, for any 3×3 matrix B we can express the product BA as

$$BA = B[\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{0}] = [B\mathbf{c}_1 \quad B\mathbf{c}_2 \quad \mathbf{0}] \qquad \textbf{[Formula (6) of Section 1.3]}$$

The column of zeros shows that $BA \neq I$ and hence that A is singular.

As in Example 6, we will frequently denote a zero matrix with one row or one column by a boldface zero.

Properties of Inverses

It is reasonable to ask whether an invertible matrix can have more than one inverse. The next theorem shows that the answer is no—*an invertible matrix has exactly one inverse.*

Theorem 1.4.4

If B and C are both inverses of the matrix A, then $B = C$.

Proof Since B is an inverse of A, we have $BA = I$. Multiplying both sides on the right by C gives $(BA)C = IC = C$. But it is also true that $(BA)C = B(AC) = BI = B$, so $C = B$. ■

As a consequence of this important result, we can now speak of "the" inverse of an invertible matrix. If A is invertible, then its inverse will be denoted by the symbol A^{-1}. Thus,

$$\boxed{AA^{-1} = I \quad \text{and} \quad A^{-1}A = I} \tag{1}$$

The inverse of A plays much the same role in matrix arithmetic that the reciprocal a^{-1} plays in the numerical relationships $aa^{-1} = 1$ and $a^{-1}a = 1$.

Warning The symbol A^{-1} should not be interpreted as $1/A$. Division by matrices is not a defined operation.

In the next section we will develop a method for computing the inverse of an invertible matrix of any size. For now we give the following theorem that specifies conditions under which a 2×2 matrix is invertible and provides a simple formula for its inverse.

The quantity $ad - bc$ in Theorem 1.4.5 is called the ***determinant*** of the 2×2 matrix A and is denoted by

$$\det(A) = ad - bc$$

or alternatively by

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Theorem 1.4.5

The matrix

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

is invertible if and only if $ad - bc \neq 0$, in which case the inverse is given by the formula

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \qquad (2)$$

We will omit the proof, because we will study a more general version of this theorem later. For now, you should at least confirm the validity of Formula (2) by showing that $AA^{-1} = A^{-1}A = I$.

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

FIGURE 1.4.1

Remark **Figure 1.4.1** illustrates that the determinant of a 2×2 matrix A is the product of the entries on its main diagonal minus the product of the entries *off* its main diagonal.

Historical Note

The formula for A^{-1} given in Theorem 1.4.5 first appeared (in a more general form) in Arthur Cayley's 1858 *Memoir on the Theory of Matrices*. The more general result that Cayley discovered will be studied later.

EXAMPLE 7 | Calculating the Inverse of a 2×2 Matrix

In each part, determine whether the matrix is invertible. If so, find its inverse.

$$(a)\ A = \begin{bmatrix} 6 & 1 \\ 5 & 2 \end{bmatrix} \qquad (b)\ A = \begin{bmatrix} -1 & 2 \\ 3 & -6 \end{bmatrix}$$

Solution (*a*) The determinant of A is $\det(A) = (6)(2) - (1)(5) = 7$, which is nonzero. Thus, A is invertible, and its inverse is

$$A^{-1} = \frac{1}{7}\begin{bmatrix} 2 & -1 \\ -5 & 6 \end{bmatrix} = \begin{bmatrix} \frac{2}{7} & -\frac{1}{7} \\ -\frac{5}{7} & \frac{6}{7} \end{bmatrix}$$

We leave it for you to confirm that $AA^{-1} = A^{-1}A = I$.

Solution (*b*) The matrix is not invertible since $\det(A) = (-1)(-6) - (2)(3) = 0$.

EXAMPLE 8 | Solution of a Linear System by Matrix Inversion

A problem that arises in many applications is to solve a pair of equations of the form

$$\begin{aligned} u &= ax + by \\ v &= cx + dy \end{aligned}$$

for x and y in terms of u and v. One approach is to treat this as a linear system of two equations in the unknowns x and y and use Gauss–Jordan elimination to solve for x and y. However, because the coefficients of the unknowns are *literal* rather than *numerical*, that procedure is a little clumsy. As an alternative approach, let us replace the two equations by the single matrix equation

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

which we can rewrite as

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

If we assume that the 2×2 matrix is invertible (i.e., $ad - bc \neq 0$), then we can multiply through on the left by the inverse and rewrite the equation as

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

which simplifies to

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

Using Theorem 1.4.5, we can rewrite this equation as

$$\frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

from which we obtain

$$x = \frac{du - bv}{ad - bc}, \quad y = \frac{av - cu}{ad - bc}$$

The next theorem is concerned with inverses of matrix products.

Theorem 1.4.6

If A and B are invertible matrices with the same size, then AB is invertible and

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

Proof We can establish the invertibility and obtain the stated formula at the same time by showing that

$$(AB)(B^{-1}A^{-1}) = (B^{-1}A^{-1})(AB) = I$$

But

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

and similarly, $(B^{-1}A^{-1})(AB) = I$. ■

Although we will not prove it, this result can be extended to three or more factors:

A product of any number of invertible matrices is invertible, and the inverse of the product is the product of the inverses in the reverse order.

EXAMPLE 9 | The Inverse of a Product

Consider the matrices

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

If a product of matrices is singular, then at least one of the factors must be singular. Why?

We leave it for you to show that

$$AB = \begin{bmatrix} 7 & 6 \\ 9 & 8 \end{bmatrix}, \quad (AB)^{-1} = \begin{bmatrix} 4 & -3 \\ -\frac{9}{2} & \frac{7}{2} \end{bmatrix}$$

and also that

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

Thus, $(AB)^{-1} = B^{-1}A^{-1}$ as guaranteed by Theorem 1.4.6.

Powers of a Matrix

If A is a *square* matrix, then we define the nonnegative integer powers of A to be

$$\boxed{A^0 = I \quad \text{and} \quad A^n = AA\cdots A \quad [n \text{ factors}]}$$

and if A is invertible, then we define the negative integer powers of A to be

$$\boxed{A^{-n} = (A^{-1})^n = A^{-1}A^{-1}\cdots A^{-1} \quad [n \text{ factors}]}$$

Because these definitions parallel those for real numbers, the usual laws of nonnegative exponents hold; for example,

$$\boxed{A^rA^s = A^{r+s} \quad \text{and} \quad (A^r)^s = A^{rs}}$$

In addition, we have the following properties of negative exponents.

Theorem 1.4.7

If A is invertible and n is a nonnegative integer, then:

(a) A^{-1} is invertible and $(A^{-1})^{-1} = A$.

(b) A^n is invertible and $(A^n)^{-1} = A^{-n} = (A^{-1})^n$.

(c) kA is invertible for any nonzero scalar k, and $(kA)^{-1} = k^{-1}A^{-1}$.

We will prove part (c) and leave the proofs of parts (a) and (b) as exercises.

Proof (c) Properties (m) and (l) of Theorem 1.4.1 imply that

$$(kA)(k^{-1}A^{-1}) = k^{-1}(kA)A^{-1} = (k^{-1}k)AA^{-1} = (1)I = I$$

and similarly, $(k^{-1}A^{-1})(kA) = I$. Thus, kA is invertible and $(kA)^{-1} = k^{-1}A^{-1}$. ■

EXAMPLE 10 | Properties of Exponents

Let A and A^{-1} be the matrices in Example 9; that is,

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \quad \text{and} \quad A^{-1} = \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix}$$

Then

$$A^{-3} = (A^{-1})^3 = \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 41 & -30 \\ -15 & 11 \end{bmatrix}$$

Also,

$$A^3 = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 11 & 30 \\ 15 & 41 \end{bmatrix}$$

so, as expected from Theorem 1.4.7(b),

$$(A^3)^{-1} = \frac{1}{(11)(41) - (30)(15)}\begin{bmatrix} 41 & -30 \\ -15 & 11 \end{bmatrix} = \begin{bmatrix} 41 & -30 \\ -15 & 11 \end{bmatrix} = (A^{-1})^3$$

EXAMPLE 11 | The Square of a Matrix Sum

In real arithmetic, where we have a commutative law for multiplication, we can write

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

However, in matrix arithmetic, where we have no commutative law for multiplication, the best we can do is to write

$$(A+B)^2 = A^2 + AB + BA + B^2$$

It is only in the special case where A and B *commute* (i.e., $AB = BA$) that we can go a step further and write

$$(A+B)^2 = A^2 + 2AB + B^2$$

Matrix Polynomials

If A is a square matrix, say $n \times n$, and if

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$$

is any polynomial, then we define the $n \times n$ matrix $p(A)$ to be

$$p(A) = a_0I + a_1A + a_2A^2 + \cdots + a_mA^m \tag{3}$$

where I is the $n \times n$ identity matrix; that is, $p(A)$ is obtained by substituting A for x and replacing the constant term a_0 by the matrix a_0I. An expression of form (3) is called a ***matrix polynomial in A***.

EXAMPLE 12 | A Matrix Polynomial

Find $p(A)$ for

$$p(x) = x^2 - 2x - 5 \quad \text{and} \quad A = \begin{bmatrix} -1 & 2 \\ 1 & 3 \end{bmatrix}$$

Solution

$$\begin{aligned} p(A) &= A^2 - 2A - 5I \\ &= \begin{bmatrix} -1 & 2 \\ 1 & 3 \end{bmatrix}^2 - 2\begin{bmatrix} -1 & 2 \\ 1 & 3 \end{bmatrix} - 5\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 4 \\ 2 & 11 \end{bmatrix} - \begin{bmatrix} -2 & 4 \\ 2 & 6 \end{bmatrix} - \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

or more briefly, $p(A) = 0$.

Remark It follows from the fact that $A^rA^s = A^{r+s} = A^{s+r} = A^sA^r$ that powers of a square matrix commute, and since a matrix polynomial in A is built up from powers of A, any two matrix polynomials in A also commute; that is, for any polynomials p_1 and p_2 we have

$$p_1(A)p_2(A) = p_2(A)p_1(A) \tag{4}$$

Properties of the Transpose

The following theorem lists the main properties of the transpose.

Theorem 1.4.8

If the sizes of the matrices are such that the stated operations can be performed, then:

(a) $(A^T)^T = A$

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

(c) $(A - B)^T = A^T - B^T$

(d) $(kA)^T = kA^T$

(e) $(AB)^T = B^TA^T$

If you keep in mind that transposing a matrix interchanges its rows and columns, then you should have little trouble visualizing the results in parts (a)–(d). For example, part (a) states the obvious fact that interchanging rows and columns twice leaves a matrix unchanged; and part (b) states that adding two matrices and then interchanging the rows and columns produces the same result as interchanging the rows and columns before adding. We will omit the formal proofs. Part (e) is less obvious, but for brevity we will omit its proof as well. The result in that part can be extended to three or more factors and restated as:

The transpose of a product of any number of matrices is the product of the transposes in the reverse order.

The following theorem establishes a relationship between the inverse of a matrix and the inverse of its transpose.

Theorem 1.4.9

If A is an invertible matrix, then A^T is also invertible and

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

Proof We can establish the invertibility and obtain the formula at the same time by showing that

$$A^T(A^{-1})^T = (A^{-1})^TA^T = I$$

But from part (e) of Theorem 1.4.8 and the fact that $I^T = I$, we have

$$A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I$$
$$(A^{-1})^TA^T = (AA^{-1})^T = I^T = I$$

which completes the proof. ■

EXAMPLE 13 | Inverse of a Transpose

Consider a general 2×2 invertible matrix and its transpose:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{and} \quad A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

Since A is invertible, its determinant $ad - bc$ is nonzero. But the determinant of A^T is also $ad - bc$ (verify), so A^T is also invertible. It follows from Theorem 1.4.5 that

$$(A^T)^{-1} = \begin{bmatrix} \dfrac{d}{ad-bc} & -\dfrac{c}{ad-bc} \\ -\dfrac{b}{ad-bc} & \dfrac{a}{ad-bc} \end{bmatrix}$$

which is the same matrix that results if A^{-1} is transposed (verify). Thus,

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

as guaranteed by Theorem 1.4.9.

Exercise Set 1.4

In Exercises 1–2, verify that the following matrices and scalars satisfy the stated properties of Theorem 1.4.1.

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

$$C = \begin{bmatrix} 4 & 1 \\ -3 & -2 \end{bmatrix}, \quad a = 4, \quad b = -7$$

1. **a.** The associative law for matrix addition.

b. The associative law for matrix multiplication.

c. The left distributive law.

d. $(a + b)C = aC + bC$

2. **a.** $a(BC) = (aB)C = B(aC)$

b. $A(B - C) = AB - AC$ **c.** $(B + C)A = BA + CA$

d. $a(bC) = (ab)C$

In Exercises 3–4, verify that the matrices and scalars in Exercise 1 satisfy the stated properties.

3. **a.** $(A^T)^T = A$ **b.** $(AB)^T = B^T A^T$

4. **a.** $(A + B)^T = A^T + B^T$ **b.** $(aC)^T = aC^T$

In Exercises 5–8, use Theorem 1.4.5 to compute the inverse of the matrix.

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

7. $C = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ **8.** $D = \begin{bmatrix} 6 & 4 \\ -2 & -1 \end{bmatrix}$

9. Find the inverse of

$$\begin{bmatrix} \frac{1}{2}(e^x + e^{-x}) & \frac{1}{2}(e^x - e^{-x}) \\ \frac{1}{2}(e^x - e^{-x}) & \frac{1}{2}(e^x + e^{-x}) \end{bmatrix}$$

10. Find the inverse of

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

In Exercises 11–14, verify that the equations are valid for the matrices in Exercises 5–8.

11. $(A^T)^{-1} = (A^{-1})^T$ **12.** $(A^{-1})^{-1} = A$

13. $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$ **14.** $(ABC)^T = C^T B^T A^T$

In Exercises 15–18, use the given information to find A.

15. $(7A)^{-1} = \begin{bmatrix} -3 & 7 \\ 1 & -2 \end{bmatrix}$ **16.** $(5A^T)^{-1} = \begin{bmatrix} -3 & -1 \\ 5 & 2 \end{bmatrix}$

17. $(I + 2A)^{-1} = \begin{bmatrix} -1 & 2 \\ 4 & 5 \end{bmatrix}$ **18.** $A^{-1} = \begin{bmatrix} 2 & -1 \\ 3 & 5 \end{bmatrix}$

In Exercises 19–20, compute the following using the given matrix A.

a. A^3 **b.** A^{-3} **c.** $A^2 - 2A + I$

19. $A = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}$ **20.** $A = \begin{bmatrix} 2 & 0 \\ 4 & 1 \end{bmatrix}$

In Exercises 21–22, compute $p(A)$ for the given matrix A and the following polynomials.

a. $p(x) = x - 2$

b. $p(x) = 2x^2 - x + 1$

c. $p(x) = x^3 - 2x + 1$

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

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

In Exercises 23–24, let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

23. Find all values of a, b, c, and d (if any) for which the matrices A and B commute.

24. Find all values of a, b, c, and d (if any) for which the matrices A and C commute.

In Exercises 25–28, use the method of Example 8 to find the unique solution of the given linear system.

25. $\begin{aligned} 3x_1 - 2x_2 &= -1 \\ 4x_1 + 5x_2 &= 3 \end{aligned}$

26. $\begin{aligned} -x_1 + 5x_2 &= 4 \\ -x_1 - 3x_2 &= 1 \end{aligned}$

27. $\begin{aligned} 6x_1 + x_2 &= 0 \\ 4x_1 - 3x_2 &= -2 \end{aligned}$

28. $\begin{aligned} 2x_1 - 2x_2 &= 4 \\ x_1 + 4x_2 &= 4 \end{aligned}$

If a polynomial $p(x)$ can be factored as a product of lower degree polynomials, say

$$p(x) = p_1(x)p_2(x)$$

and if A is a square matrix, then it can be proved that

$$p(A) = p_1(A)p_2(A)$$

In Exercises 29–30, verify this statement for the stated matrix A and polynomials

$$p(x) = x^2 - 9, \quad p_1(x) = x + 3, \quad p_2(x) = x - 3$$

29. The matrix A in Exercise 21.

30. An arbitrary square matrix A.

31. a. Give an example of two 2×2 matrices such that

$$(A + B)(A - B) \neq A^2 - B^2$$

b. State a valid formula for multiplying out

$$(A + B)(A - B)$$

c. What condition can you impose on A and B that will allow you to write $(A + B)(A - B) = A^2 - B^2$?

32. The numerical equation $a^2 = 1$ has exactly two solutions. Find at least eight solutions of the matrix equation $A^2 = I_3$. [*Hint:* Look for solutions in which all entries off the main diagonal are zero.]

33. a. Show that if a square matrix A satisfies the equation $A^2 + 2A + I = 0$, then A must be invertible. What is the inverse?

b. Show that if $p(x)$ is a polynomial with a nonzero constant term, and if A is a square matrix for which $p(A) = 0$, then A is invertible.

34. Is it possible for A^3 to be an identity matrix without A being invertible? Explain.

35. Can a matrix with a row of zeros or a column of zeros have an inverse? Explain.

36. Can a matrix with two identical rows or two identical columns have an inverse? Explain.

In Exercises 37–38, determine whether A is invertible, and if so, find the inverse. [Hint: *Solve $AX = I$ for X by equating corresponding entries on the two sides.*]

37. $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$

38. $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$

In Exercises 39–40, simplify the expression assuming that A, B, C, and D are invertible.

39. $(AB)^{-1}(AC^{-1})(D^{-1}C^{-1})^{-1}D^{-1}$

40. $(AC^{-1})^{-1}(AC^{-1})(AC^{-1})^{-1}AD^{-1}$

41. Show that if R is a $1 \times n$ matrix and C is an $n \times 1$ matrix, then $RC = \operatorname{tr}(CR)$.

42. If A is a square matrix and n is a positive integer, is it true that $(A^n)^T = (A^T)^n$? Justify your answer.

43. a. Show that if A is invertible and $AB = AC$, then $B = C$.

b. Explain why part (a) and Example 3 do not contradict one another.

44. Show that if A is invertible and k is any nonzero scalar, then $(kA)^n = k^n A^n$ for all integer values of n.

45. a. Show that if A, B, and $A + B$ are invertible matrices with the same size, then

$$A(A^{-1} + B^{-1})B(A + B)^{-1} = I$$

b. What does the result in part (a) tell you about the matrix $A^{-1} + B^{-1}$?

46. A square matrix A is said to be ***idempotent*** if $A^2 = A$.

a. Show that if A is idempotent, then so is $I - A$.

b. Show that if A is idempotent, then $2A - I$ is invertible and is its own inverse.

47. Show that if A is a square matrix such that $A^k = 0$ for some positive integer k, then the matrix $I - A$ is invertible and

$$(I - A)^{-1} = I + A + A^2 + \cdots + A^{k-1}$$

48. Show that the matrix

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

satisfies the equation

$$A^2 - (a + d)A + (ad - bc)I = 0$$

49. Assuming that all matrices are $n \times n$ and invertible, solve for D.

$$C^T B^{-1} A^2 B A C^{-1} D A^{-2} B^T C^{-2} = C^T$$

50. Assuming that all matrices are $n \times n$ and invertible, solve for D.

$$A B C^T D B A^T C = A B^T$$

Working with Proofs

In Exercises 51–58, prove the stated result.

51. Theorem 1.4.1(a)

52. Theorem 1.4.1(b)

53. Theorem 1.4.1(f)

54. Theorem 1.4.1(c)

55. Theorem 1.4.2(c)

56. Theorem 1.4.2(b)

57. Theorem 1.4.8(d)

58. Theorem 1.4.8(e)

True-False Exercises

TF. In parts **(a)–(k)** determine whether the statement is true or false, and justify your answer.

a. Two $n \times n$ matrices, A and B, are inverses of one another if and only if $AB = BA = 0$.

b. For all square matrices A and B of the same size, it is true that $(A+B)^2 = A^2 + 2AB + B^2$.

c. For all square matrices A and B of the same size, it is true that $A^2 - B^2 = (A-B)(A+B)$.

d. If A and B are invertible matrices of the same size, then AB is invertible and $(AB)^{-1} = A^{-1}B^{-1}$.

e. If A and B are matrices such that AB is defined, then it is true that $(AB)^T = A^T B^T$.

f. The matrix

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

is invertible if and only if $ad - bc \neq 0$.

g. If A and B are matrices of the same size and k is a constant, then $(kA + B)^T = kA^T + B^T$.

h. If A is an invertible matrix, then so is A^T.

i. If $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$ and I is an identity matrix, then $p(I) = a_0 + a_1 + a_2 + \cdots + a_m$.

j. A square matrix containing a row or column of zeros cannot be invertible.

k. The sum of two invertible matrices of the same size must be invertible.

Working with Technology

T1. Let A be the matrix

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

Discuss the behavior of A^k as k increases indefinitely, that is, as $k \to \infty$.

T2. In each part use your technology utility to make a conjecture about the form of A^n for positive integer powers of n.

a. $A = \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix}$ **b.** $A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$

T3. The ***Fibonacci sequence*** (named for the Italian mathematician Leonardo Fibonacci 1170–1250) is

$$0,\ 1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ 55,\ 89,\ 144, \ldots$$

the terms of which are commonly denoted as

$$F_0,\ F_1,\ F_2,\ F_3, \ldots,\ F_n, \ldots$$

After the initial terms $F_0 = 0$ and $F_1 = 1$, each term is the sum of the previous two; that is,

$$F_n = F_{n-1} + F_{n-2}$$

Confirm that if

$$Q = \begin{bmatrix} F_2 & F_1 \\ F_1 & F_0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

then

$$Q^n = \begin{bmatrix} F_{n+1} & F_n \\ F_n & F_0 \end{bmatrix}$$

1.5 Elementary Matrices and a Method for Finding A^{-1}

In this section we will develop an algorithm for finding the inverse of a matrix, and we will discuss some of the basic properties of invertible matrices.

Elementary Matrices

In Section 1.1 we defined three elementary row operations on a matrix A:

1. Multiply a row by a nonzero constant c.
2. Interchange two rows.
3. Add a constant c times one row to another.

It should be evident that if we let B be the matrix that results from A by performing one of the operations in this list, then the matrix A can be recovered from B by performing the corresponding operation in the following list:

1. Multiply the same row by $1/c$.
2. Interchange the same two rows.
3. If B resulted by adding c times row r_i of A to row r_j, then add $-c$ times r_j to r_i.

It follows that if B is obtained from A by performing a sequence of elementary row operations, then there is a second sequence of elementary row operations, which when applied to B recovers A. Accordingly, we make the following definition.

Definition 1

Matrices A and B are said to be ***row equivalent*** if either (hence each) can be obtained from the other by a sequence of elementary row operations.

Our next goal is to show how matrix multiplication can be used to carry out an elementary row operation.

Definition 2

A matrix E is called an ***elementary matrix*** if it can be obtained from an identity matrix by performing a *single* elementary row operation.

EXAMPLE 1 | Elementary Matrices and Row Operations

Listed below are four elementary matrices and the operations that produce them.

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

Multiply the second row of I_2 by -3.

Interchange the second and fourth rows of I_4.

Add 3 times the third row of I_3 to the first row.

Multiply the first row of I_3 by 1.

The following theorem, whose proof is left as an exercise, shows that when a matrix A is multiplied on the *left* by an elementary matrix E, the effect is to perform an elementary row operation on A.

Theorem 1.5.1

Row Operations by Matrix Multiplication

If the elementary matrix E results from performing a certain row operation on I_m and if A is an $m \times n$ matrix, then the product EA is the matrix that results when this same row operation is performed on A.

Theorem 1.5.1 will be a useful tool for developing new results about matrices, but as a practical matter it is usually preferable to perform row operations directly.

EXAMPLE 2 | Using Elementary Matrices

Consider the matrix

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

and consider the elementary matrix

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

which results from adding 3 times the first row of I_3 to the third row. The product EA is

$$EA = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 4 & 4 & 10 & 9 \end{bmatrix}$$

which is precisely the matrix that results when we add 3 times the first row of A to the third row.

We know from the discussion at the beginning of this section that if E is an elementary matrix that results from performing an elementary row operation on an identity matrix I, then there is a second elementary row operation, which when applied to E produces I back again. **Table 1** lists these operations. The operations on the right side of the table are called the ***inverse operations*** of the corresponding operations on the left.

TABLE 1

Row Operation on I That Produces E	**Row Operation on E That Reproduces I**
Multiply row i by $c \neq 0$	Multiply row i by $1/c$
Interchange rows i and j	Interchange rows i and j
Add c times row i to row j	Add $-c$ times row i to row j

EXAMPLE 3 | Row Operations and Inverse Row Operations

In each of the following, an elementary row operation is applied to the 2×2 identity matrix to obtain an elementary matrix E, then E is restored to the identity matrix by applying the inverse row operation.

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

Multiply the second row by 7.

Multiply the second row by $\frac{1}{7}$.

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

Interchange the first and second rows. Interchange the first and second rows.

$$\begin{bmatrix}1&0\\0&1\end{bmatrix} \longrightarrow \begin{bmatrix}1&5\\0&1\end{bmatrix} \longrightarrow \begin{bmatrix}1&0\\0&1\end{bmatrix}$$

Add 5 times the second row to the first. Add -5 times the second row to the first.

The next theorem is a key result about invertibility of elementary matrices. It will be a building block for many results that follow.

Theorem 1.5.2

Every elementary matrix is invertible, and the inverse is also an elementary matrix.

Proof If E is an elementary matrix, then E results by performing some row operation on I. Let E_0 be the matrix that results when the inverse of that operation is performed on I. Applying Theorem 1.5.1 and using the fact that inverse row operations cancel the effect of each other, it follows that

$$E_0E = I \quad \text{and} \quad EE_0 = I$$

Thus, the elementary matrix E_0 is the inverse of E. ■

Equivalence Theorem

One of our objectives as we progress through this text is to show how seemingly diverse ideas in linear algebra are related. The following theorem, which relates results we have obtained about invertibility of matrices, homogeneous linear systems, reduced row echelon forms, and elementary matrices, is our first step in that direction. As we study new topics, more statements will be added to this theorem.

Theorem 1.5.3

Equivalent Statements

If A is an $n \times n$ matrix, then the following statements are equivalent, that is, all true or all false.

(a) A is invertible.

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

(c) The reduced row echelon form of A is I_n.

(d) A is expressible as a product of elementary matrices.

Proof We will prove the equivalence by establishing the chain of implications: $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (d) \Rightarrow (a)$.

The following figure illustrates that the sequence of implications

$$(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (d) \Rightarrow (a)$$

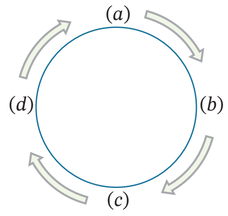
implies that

$$(d) \Rightarrow (c) \Rightarrow (b) \Rightarrow (a)$$

and hence that

$$(a) \Leftrightarrow (b) \Leftrightarrow (c) \Leftrightarrow (d)$$

(see Appendix A).



(*a*) ⇒ (*b*) Assume A is invertible and let $\mathbf{x}_0$ be any solution of $A\mathbf{x} = \mathbf{0}$. Multiplying both sides of this equation by A^{-1} gives

$$(A^{-1}A)\mathbf{x}_0 = A^{-1}\mathbf{0}$$

from which it follows that $\mathbf{x}_0 = \mathbf{0}$, so $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(*b*) ⇒ (*c*) Let $A\mathbf{x} = \mathbf{0}$ be the matrix form of the system

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + \cdots + & a_{1n}x_n & = 0 \\ a_{21}x_1 & + & a_{22}x_2 & + \cdots + & a_{2n}x_n & = 0 \\ \vdots & & \vdots & & \vdots & \vdots \\ a_{n1}x_1 & + & a_{n2}x_2 & + \cdots + & a_{nn}x_n & = 0 \end{array} \tag{1}$$

and assume that the system has only the trivial solution. If we solve by Gauss–Jordan elimination, then the system of equations corresponding to the reduced row echelon form of the augmented matrix will be

$$\begin{array}{cccc} x_1 & & & = 0 \\ & x_2 & & = 0 \\ & & \ddots & \\ & & & x_n = 0 \end{array} \tag{2}$$

Thus, the augmented matrix

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & 0 \end{bmatrix}$$

for (1) can be reduced to the augmented matrix

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

for (2) by a sequence of elementary row operations. If we disregard the last column (all zeros) in each of these matrices, we can conclude that the reduced row echelon form of A is I_n.

(*c*) ⇒ (*d*) Assume that the reduced row echelon form of A is I_n, so that A can be reduced to I_n by a finite sequence of elementary row operations. By Theorem 1.5.1, each of these operations can be accomplished by multiplying on the left by an appropriate elementary matrix. Thus we can find elementary matrices $E_1, E_2, \ldots, E_k$ such that

$$E_k \cdots E_2E_1A = I_n \tag{3}$$

By Theorem 1.5.2, $E_1, E_2, \ldots, E_k$ are invertible. Multiplying both sides of Equation (3) on the left successively by $E_k^{-1}, \ldots, E_2^{-1}, E_1^{-1}$ we obtain

$$A = E_1^{-1}E_2^{-1} \cdots E_k^{-1}I_n = E_1^{-1}E_2^{-1} \cdots E_k^{-1} \tag{4}$$

By Theorem 1.5.2, this equation expresses A as a product of elementary matrices.

(*d*) ⇒ (*a*) If A is a product of elementary matrices, then from Theorems 1.4.6 and 1.5.2, the matrix A is a product of invertible matrices and hence is invertible. ■

A Method for Inverting Matrices

As a first application of Theorem 1.5.3, we will develop a procedure (or algorithm) that can be used to tell whether a given matrix is invertible, and if so, produce its inverse. To derive this algorithm, assume for the moment, that A is an invertible $n \times n$ matrix. In Equation (3), the elementary matrices execute a sequence of row operations that reduce A to I_n. If we multiply both sides of this equation on the right by A^{-1} and simplify, we obtain

$$A^{-1} = E_k \cdots E_2 E_1 I_n$$

But this equation tells us that *the same sequence of row operations that reduces A to I_n will transform I_n to A^{-1}*. Thus, we have established the following result.

> **Inversion Algorithm** To find the inverse of an invertible matrix A, find a sequence of elementary row operations that reduces A to the identity and then perform that same sequence of operations on I_n to obtain A^{-1}.

A simple method for carrying out this procedure is given in the following example.

EXAMPLE 4 | Using Row Operations to Find A^{-1}

Find the inverse of

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

Solution We want to reduce A to the identity matrix by row operations and simultaneously apply these operations to I to produce A^{-1}. To accomplish this we will adjoin the identity matrix to the right side of A, thereby producing a partitioned matrix of the form

$$[A \mid I]$$

Then we will apply row operations to this matrix until the left side is reduced to I; these operations will convert the right side to A^{-1}, so the final matrix will have the form

$$[I \mid A^{-1}]$$

The computations are as follows:

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

$$\left[\begin{array}{rrr|rrr} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & -2 & 5 & -1 & 0 & 1 \end{array}\right] \longleftarrow$$
We added -2 times the first row to the second and -1 times the first row to the third.

$$\left[\begin{array}{rrr|rrr} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & -1 & -5 & 2 & 1 \end{array}\right] \longleftarrow$$
We added 2 times the second row to the third.

$$\left[\begin{array}{rrr|rrr} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right] \longleftarrow$$
We multiplied the third row by -1.

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & -14 & 6 & 3 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right]$$ ← We added 3 times the third row to the second and −3 times the third row to the first.

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -40 & 16 & 9 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right]$$ ← We added −2 times the second row to the first.

Thus,

$$A^{-1} = \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix}$$

Often it will not be known in advance if a given $n \times n$ matrix A is invertible. However, if it is not, then by parts (a) and (c) of Theorem 1.5.3 it will be impossible to reduce A to I_n by elementary row operations. This will be signaled by a row of zeros appearing on the *left side* of the partition at some stage of the inversion algorithm. If this occurs, then you can stop the computations and conclude that A is not invertible.

EXAMPLE 5 | Showing That a Matrix Is Not Invertible

Consider the matrix

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

Applying the procedure of Example 4 yields

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

$$\left[\begin{array}{ccc|ccc} 1 & 6 & 4 & 1 & 0 & 0 \\ 0 & -8 & -9 & -2 & 1 & 0 \\ 0 & 8 & 9 & 1 & 0 & 1 \end{array}\right]$$ ← We added −2 times the first row to the second and added the first row to the third.

$$\left[\begin{array}{ccc|ccc} 1 & 6 & 4 & 1 & 0 & 0 \\ 0 & -8 & -9 & -2 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 \end{array}\right]$$ ← We added the second row to the third.

Since we have obtained a row of zeros on the left side, A is not invertible.

EXAMPLE 6 | Analyzing Homogeneous Systems

Use Theorem 1.5.3 to determine whether the given homogeneous system has nontrivial solutions.

(a)
$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 0 \\ 2x_1 + 5x_2 + 3x_3 &= 0 \\ x_1 + 8x_3 &= 0 \end{aligned}$$

(b)
$$\begin{aligned} x_1 + 6x_2 + 4x_3 &= 0 \\ 2x_1 + 4x_2 - x_3 &= 0 \\ -x_1 + 2x_2 + 5x_3 &= 0 \end{aligned}$$

Solution From parts (a) and (b) of Theorem 1.5.3 a homogeneous linear system has only the trivial solution if and only if its coefficient matrix is invertible. From Examples 4 and 5 the coefficient matrix of system (a) is invertible and that of system (b) is not. Thus, system (a) has only the trivial solution while system (b) has nontrivial solutions.

Exercise Set 1.5

In Exercises 1–2, determine whether the given matrix is elementary.

1. **a.** $\begin{bmatrix} 1 & 0 \\ -5 & 1 \end{bmatrix}$ **b.** $\begin{bmatrix} -5 & 1 \\ 1 & 0 \end{bmatrix}$

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

2. **a.** $\begin{bmatrix} 1 & 0 \\ 0 & \sqrt{3} \end{bmatrix}$ **b.** $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

c. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 9 \\ 0 & 0 & 1 \end{bmatrix}$ **d.** $\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

In Exercises 3–4, find a row operation and the corresponding elementary matrix that will restore the given elementary matrix to the identity matrix.

3. **a.** $\begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix}$ **b.** $\begin{bmatrix} -7 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

c. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -5 & 0 & 1 \end{bmatrix}$ **d.** $\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

4. **a.** $\begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

c. $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$ **d.** $\begin{bmatrix} 1 & 0 & -\frac{1}{7} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

In Exercises 5–6 an elementary matrix E and a matrix A are given. Identify the row operation corresponding to E and verify that the product EA results from applying the row operation to A.

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

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

c. $E = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

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

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

c. $E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

In Exercises 7–8, use the following matrices and find an elementary matrix E that satisfies the stated equation.

$$A = \begin{bmatrix} 3 & 4 & 1 \\ 2 & -7 & -1 \\ 8 & 1 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 8 & 1 & 5 \\ 2 & -7 & -1 \\ 3 & 4 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 3 & 4 & 1 \\ 2 & -7 & -1 \\ 2 & -7 & 3 \end{bmatrix}, \quad D = \begin{bmatrix} 8 & 1 & 5 \\ -6 & 21 & 3 \\ 3 & 4 & 1 \end{bmatrix}$$

$$F = \begin{bmatrix} 8 & 1 & 5 \\ 8 & 1 & 1 \\ 3 & 4 & 1 \end{bmatrix}$$

7. **a.** $EA = B$ **b.** $EB = A$

c. $EA = C$ **d.** $EC = A$

8. **a.** $EB = D$ **b.** $ED = B$

c. $EB = F$ **d.** $EF = B$

In Exercises 9–10, first use Theorem 1.4.5 and then use the inversion algorithm to find A^{-1}, if it exists.

9. **a.** $A = \begin{bmatrix} 1 & 4 \\ 2 & 7 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 2 & -4 \\ -4 & 8 \end{bmatrix}$

10. **a.** $A = \begin{bmatrix} 1 & -5 \\ 3 & -16 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 6 & 4 \\ -3 & -2 \end{bmatrix}$

In Exercises 11–12, use the inversion algorithm to find the inverse of the matrix (if the inverse exists).

11. **a.** $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}$ **b.** $\begin{bmatrix} -1 & 3 & -4 \\ 2 & 4 & 1 \\ -4 & 2 & -9 \end{bmatrix}$

12. **a.** $\begin{bmatrix} \frac{1}{5} & \frac{1}{5} & -\frac{2}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{10} \\ \frac{1}{5} & -\frac{4}{5} & \frac{1}{10} \end{bmatrix}$ **b.** $\begin{bmatrix} \frac{1}{5} & \frac{1}{5} & -\frac{2}{5} \\ \frac{2}{5} & -\frac{3}{5} & -\frac{3}{10} \\ \frac{1}{5} & -\frac{4}{5} & \frac{1}{10} \end{bmatrix}$

In Exercises 13–18, use the inversion algorithm to find the inverse of the matrix (if the inverse exists).

13. $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ **14.** $\begin{bmatrix} \sqrt{2} & 3\sqrt{2} & 0 \\ -4\sqrt{2} & \sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$

15. $\begin{bmatrix} 2 & 6 & 6 \\ 2 & 7 & 6 \\ 2 & 7 & 7 \end{bmatrix}$ **16.** $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 5 & 0 \\ 1 & 3 & 5 & 7 \end{bmatrix}$

17. $\begin{bmatrix} 2 & -4 & 0 & 0 \\ 1 & 2 & 12 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & -1 & -4 & -5 \end{bmatrix}$ **18.** $\begin{bmatrix} 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & -1 & 3 & 0 \\ 2 & 1 & 5 & -3 \end{bmatrix}$

In Exercises 19–20, find the inverse of each of the following 4×4 matrices, where k_1, k_2, k_3, k_4, and k are all nonzero.

19. a. $\begin{bmatrix} k_1 & 0 & 0 & 0 \\ 0 & k_2 & 0 & 0 \\ 0 & 0 & k_3 & 0 \\ 0 & 0 & 0 & k_4 \end{bmatrix}$ **b.** $\begin{bmatrix} k & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & k & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

20. a. $\begin{bmatrix} 0 & 0 & 0 & k_1 \\ 0 & 0 & k_2 & 0 \\ 0 & k_3 & 0 & 0 \\ k_4 & 0 & 0 & 0 \end{bmatrix}$ **b.** $\begin{bmatrix} k & 0 & 0 & 0 \\ 1 & k & 0 & 0 \\ 0 & 1 & k & 0 \\ 0 & 0 & 1 & k \end{bmatrix}$

In Exercises 21–22, find all values of c, if any, for which the given matrix is invertible.

21. $\begin{bmatrix} c & c & c \\ 1 & c & c \\ 1 & 1 & c \end{bmatrix}$ **22.** $\begin{bmatrix} c & 1 & 0 \\ 1 & c & 1 \\ 0 & 1 & c \end{bmatrix}$

In Exercises 23–26, express the matrix and its inverse as products of elementary matrices.

23. $\begin{bmatrix} -3 & 1 \\ 2 & 2 \end{bmatrix}$ **24.** $\begin{bmatrix} 1 & 0 \\ -5 & 2 \end{bmatrix}$

25. $\begin{bmatrix} 1 & 0 & -2 \\ 0 & 4 & 3 \\ 0 & 0 & 1 \end{bmatrix}$ **26.** $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$

In Exercises 27–28, show that the matrices A and B are row equivalent by finding a sequence of elementary row operations that produces B from A, and then use that result to find a matrix C such that $CA = B$.

27. $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 1 \\ 2 & 1 & 9 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 & 5 \\ 0 & 2 & -2 \\ 1 & 1 & 4 \end{bmatrix}$

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

29. Show that if

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

is an elementary matrix, then at least one entry in the third row must be zero.

30. Show that

$$A = \begin{bmatrix} 0 & a & 0 & 0 & 0 \\ b & 0 & c & 0 & 0 \\ 0 & d & 0 & e & 0 \\ 0 & 0 & f & 0 & g \\ 0 & 0 & 0 & h & 0 \end{bmatrix}$$

is not invertible for any values of the entries.

Working with Proofs

31. Prove that if A and B are $m \times n$ matrices, then A and B are row equivalent if and only if A and B have the same reduced row echelon form.

32. Prove that if A is an invertible matrix and B is row equivalent to A, then B is also invertible.

33. Prove that if B is obtained from A by performing a sequence of elementary row operations, then there is a second sequence of elementary row operations, which when applied to B recovers A.

True-False Exercises

TF. In parts (a)–(g) determine whether the statement is true or false, and justify your answer.

a. The product of two elementary matrices of the same size must be an elementary matrix.

b. Every elementary matrix is invertible.

c. If A and B are row equivalent, and if B and C are row equivalent, then A and C are row equivalent.

d. If A is an $n \times n$ matrix that is not invertible, then the linear system $A\mathbf{x} = 0$ has infinitely many solutions.

e. If A is an $n \times n$ matrix that is not invertible, then the matrix obtained by interchanging two rows of A cannot be invertible.

f. If A is invertible and a multiple of the first row of A is added to the second row, then the resulting matrix is invertible.

g. An expression of an invertible matrix A as a product of elementary matrices is unique.

Working with Technology

T1. It can be proved that if the partitioned matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

is invertible, then its inverse is

$$\begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}$$

provided that all of the inverses on the right side exist. Use this result to find the inverse of the matrix

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

1.6 More on Linear Systems and Invertible Matrices

In this section we will show how the inverse of a matrix can be used to solve a linear system, and we will develop some more results about invertible matrices.

Number of Solutions of a Linear System

In Section 1.1 we made the statement (based on Figures 1.1.1 and 1.1.2) that every linear system either has no solutions, has exactly one solution, or has infinitely many solutions. We are now in a position to prove this fundamental result.

Theorem 1.6.1

A system of linear equations has zero, one, or infinitely many solutions. There are no other possibilities.

Proof If $A\mathbf{x} = \mathbf{b}$ is a system of linear equations, exactly one of the following is true: (a) the system has no solutions, (b) the system has exactly one solution, or (c) the system has more than one solution. The proof will be complete if we can show that the system has infinitely many solutions in case (c).

Assume that $A\mathbf{x} = \mathbf{b}$ has more than one solution, and let $\mathbf{x}_0 = \mathbf{x}_1 - \mathbf{x}_2$, where $\mathbf{x}_1$ and $\mathbf{x}_2$ are any two distinct solutions. Because $\mathbf{x}_1$ and $\mathbf{x}_2$ are distinct, the matrix $\mathbf{x}_0$ is nonzero; moreover,

$$A\mathbf{x}_0 = A(\mathbf{x}_1 - \mathbf{x}_2) = A\mathbf{x}_1 - A\mathbf{x}_2 = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

If we now let k be any scalar, then

$$\begin{aligned} A(\mathbf{x}_1 + k\mathbf{x}_0) &= A\mathbf{x}_1 + A(k\mathbf{x}_0) = A\mathbf{x}_1 + k(A\mathbf{x}_0) \\ &= \mathbf{b} + k\mathbf{0} = \mathbf{b} + \mathbf{0} = \mathbf{b} \end{aligned}$$

But this says that $\mathbf{x}_1 + k\mathbf{x}_0$ is a solution of $A\mathbf{x} = \mathbf{b}$. Since $\mathbf{x}_0$ is nonzero and there are infinitely many choices for k, the system $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions. ■

Solving Linear Systems by Matrix Inversion

Thus far we have studied two *procedures* for solving linear systems—Gauss–Jordan elimination and Gaussian elimination. The following theorem provides an actual *formula* for the solution of a linear system of n equations in n unknowns in the case where the coefficient matrix is invertible.

Theorem 1.6.2

If A is an invertible $n \times n$ matrix, then for every $n \times 1$ matrix $\mathbf{b}$, the system of equations $A\mathbf{x} = \mathbf{b}$ has exactly one solution, namely, $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof Since $A(A^{-1}\mathbf{b}) = \mathbf{b}$, it follows that $\mathbf{x} = A^{-1}\mathbf{b}$ is a solution of $A\mathbf{x} = \mathbf{b}$. To show that this is the only solution, we will assume that $\mathbf{x}_0$ is an arbitrary solution and then show that $\mathbf{x}_0$ must be the solution $A^{-1}\mathbf{b}$.

If $\mathbf{x}_0$ is any solution of $A\mathbf{x} = \mathbf{b}$, then $A\mathbf{x}_0 = \mathbf{b}$. Multiplying both sides of this equation by A^{-1}, we obtain $\mathbf{x}_0 = A^{-1}\mathbf{b}$. ■

EXAMPLE 1 | Solution of a Linear System Using A^{-1}

Consider the system of linear equations

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 5 \\ 2x_1 + 5x_2 + 3x_3 &= 3 \\ x_1 \qquad\quad + 8x_3 &= 17 \end{aligned}$$

In matrix form this system can be written as $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 5 \\ 3 \\ 17 \end{bmatrix}$$

In Example 4 of the preceding section, we showed that A is invertible and

$$A^{-1} = \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix}$$

By Theorem 1.6.2, the solution of the system is

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 17 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$$

or $x_1 = 1, x_2 = -1, x_3 = 2$.

Keep in mind that the method of Example 1 applies only when the system has as many equations as unknowns and the coefficient matrix is invertible.

Linear Systems with a Common Coefficient Matrix

Frequently, one is concerned with solving a sequence of systems

$$A\mathbf{x} = \mathbf{b}_1, \quad A\mathbf{x} = \mathbf{b}_2, \quad A\mathbf{x} = \mathbf{b}_3, \ldots, \quad A\mathbf{x} = \mathbf{b}_k$$

each of which has the same square coefficient matrix A. If A is invertible, then the solutions

$$\mathbf{x}_1 = A^{-1}\mathbf{b}_1, \quad \mathbf{x}_2 = A^{-1}\mathbf{b}_2, \quad \mathbf{x}_3 = A^{-1}\mathbf{b}_3, \ldots, \quad \mathbf{x}_k = A^{-1}\mathbf{b}_k$$

can be obtained with one matrix inversion and k matrix multiplications. An efficient way to do this is to form the partitioned matrix

$$[A \mid \mathbf{b}_1 \mid \mathbf{b}_2 \mid \cdots \mid \mathbf{b}_k] \tag{1}$$

in which the coefficient matrix A is "augmented" by all k of the matrices $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k$, and then reduce (1) to reduced row echelon form by Gauss–Jordan elimination. In this way we can solve all k systems at once. This method has the added advantage that it applies even when A is not invertible.

EXAMPLE 2 | Solving Two Linear Systems at Once

Solve the systems

(a) $$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 4 \\ 2x_1 + 5x_2 + 3x_3 &= 5 \\ x_1 \qquad\quad + 8x_3 &= 9 \end{aligned}$$

(b) $$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 1 \\ 2x_1 + 5x_2 + 3x_3 &= 6 \\ x_1 \qquad\quad + 8x_3 &= -6 \end{aligned}$$

Solution The two systems have the same coefficient matrix. If we augment this coefficient matrix with the columns of constants on the right sides of these systems, we obtain

$$\left[\begin{array}{ccc|c|c} 1 & 2 & 3 & 4 & 1 \\ 2 & 5 & 3 & 5 & 6 \\ 1 & 0 & 8 & 9 & -6 \end{array}\right]$$

Reducing this matrix to reduced row echelon form yields (verify)

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

It follows from the last two columns that the solution of system (a) is $x_1 = 1, x_2 = 0, x_3 = 1$ and the solution of system (b) is $x_1 = 2, x_2 = 1, x_3 = -1$.

Properties of Invertible Matrices

Up to now, to show that an $n \times n$ matrix A is invertible, it has been necessary to find an $n \times n$ matrix B such that

$$AB = I \quad \text{and} \quad BA = I$$

The next theorem shows that if we can produce an $n \times n$ matrix B satisfying *either* condition, then the other condition will hold automatically.

Theorem 1.6.3

Let A be a square matrix.

(a) If B is a square matrix satisfying $BA = I$, then $B = A^{-1}$.

(b) If B is a square matrix satisfying $AB = I$, then $B = A^{-1}$.

We will prove part (a) and leave part (b) as an exercise.

Proof (a) Assume that $BA = I$. If we can show that A is invertible, the proof can be completed by multiplying $BA = I$ on both sides by A^{-1} to obtain

$$BAA^{-1} = IA^{-1} \quad \text{or} \quad BI = IA^{-1} \quad \text{or} \quad B = A^{-1}$$

To show that A is invertible, it suffices to show that the system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution (see Theorem 1.5.3). Let $\mathbf{x}_0$ be any solution of this system. If we multiply both sides of $A\mathbf{x}_0 = \mathbf{0}$ on the left by B, we obtain $BA\mathbf{x}_0 = B\mathbf{0}$ or $I\mathbf{x}_0 = \mathbf{0}$ or $\mathbf{x}_0 = \mathbf{0}$. Thus, the system of equations $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. ■

Equivalence Theorem

We are now in a position to add two more statements to the four given in Theorem 1.5.3.

Theorem 1.6.4

Equivalent Statements

If A is an $n \times n$ matrix, then the following are equivalent.

(a) A is invertible.

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

(*c*) The reduced row echelon form of A is I_n.
(*d*) A is expressible as a product of elementary matrices.
(*e*) $A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.
(*f*) $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.

Proof Since we proved in Theorem 1.5.3 that (*a*), (*b*), (*c*), and (*d*) are equivalent, it will be sufficient to prove that $(a) \Rightarrow (f) \Rightarrow (e) \Rightarrow (a)$.

(*a*) ⇒ (*f*) This was already proved in Theorem 1.6.2.

(*f*) ⇒ (*e*) This is almost self-evident, for if $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$, then $A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.

(*e*) ⇒ (*a*) If the system $A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$, then, in particular, this is so for the systems

$$A\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad A\mathbf{x} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \quad A\mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be solutions of the respective systems, and let us form an $n \times n$ matrix C having these solutions as columns. Thus C has the form

$$C = [\mathbf{x}_1 \mid \mathbf{x}_2 \mid \cdots \mid \mathbf{x}_n]$$

As discussed in Section 1.3, the successive columns of the product AC will be

$$A\mathbf{x}_1, A\mathbf{x}_2, \ldots, A\mathbf{x}_n$$

[see Formula (8) of Section 1.3]. Thus,

$$AC = [A\mathbf{x}_1 \mid A\mathbf{x}_2 \mid \cdots \mid A\mathbf{x}_n] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I$$

By part (*b*) of Theorem 1.6.3, it follows that $C = A^{-1}$. Thus, A is invertible. ■

It follows from the equivalency of parts (*e*) and (*f*) that if you can show that $A\mathbf{x} = \mathbf{b}$ has *at least one* solution for every $n \times 1$ matrix $\mathbf{b}$, then you can conclude that it has *exactly one* solution for every $n \times 1$ matrix $\mathbf{b}$.

We know from earlier work that invertible matrix factors produce an invertible product. Conversely, the following theorem shows that if the product of square matrices is invertible, then the factors themselves must be invertible.

Theorem 1.6.5

Let A and B be square matrices of the same size. If AB is invertible, then A and B must also be invertible.

Proof We will show first that B is invertible by showing that the homogeneous system $B\mathbf{x} = \mathbf{0}$ has only the trivial solution. If we assume that $\mathbf{x}_0$ is any solution of this system, then

$$(AB)\mathbf{x}_0 = A(B\mathbf{x}_0) = A\mathbf{0} = \mathbf{0}$$

so $\mathbf{x}_0 = \mathbf{0}$ by parts (a) and (b) of Theorem 1.6.4 applied to the invertible matrix AB. Thus, $B\mathbf{x} = \mathbf{0}$ has only the trivial solution, which implies that B is invertible. But this in turn implies that A is invertible since A can be expressed as

$$A = A(BB^{-1}) = (AB)B^{-1}$$

which is a product of two invertible matrices. This completes the proof. ■

In our later work the following fundamental problem will occur frequently in various contexts.

A Fundamental Problem Let A be a fixed $m \times n$ matrix. Find all $m \times 1$ matrices $\mathbf{b}$ such that the system of equations $A\mathbf{x} = \mathbf{b}$ is consistent.

If A is an invertible matrix, Theorem 1.6.2 completely solves this problem by asserting that for *every* $m \times 1$ matrix $\mathbf{b}$, the linear system $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$. If A is not square, or if A is square but not invertible, then Theorem 1.6.2 does not apply. In these cases $\mathbf{b}$ must usually satisfy certain conditions in order for $A\mathbf{x} = \mathbf{b}$ to be consistent. The following example illustrates how the methods of Section 1.2 can be used to determine such conditions.

EXAMPLE 3 | Determining Consistency by Elimination

What conditions must b_1, b_2, and b_3 satisfy in order for the system of equations

$$\begin{aligned} x_1 + x_2 + 2x_3 &= b_1 \\ x_1 + x_3 &= b_2 \\ 2x_1 + x_2 + 3x_3 &= b_3 \end{aligned}$$

to be consistent?

Solution The augmented matrix is

$$\begin{bmatrix} 1 & 1 & 2 & b_1 \\ 1 & 0 & 1 & b_2 \\ 2 & 1 & 3 & b_3 \end{bmatrix}$$

which can be reduced to row echelon form as follows:

$$\begin{bmatrix} 1 & 1 & 2 & b_1 \\ 0 & -1 & -1 & b_2 - b_1 \\ 0 & -1 & -1 & b_3 - 2b_1 \end{bmatrix}$$

-1 times the first row was added to the second and -2 times the first row was added to the third.

$$\begin{bmatrix} 1 & 1 & 2 & b_1 \\ 0 & 1 & 1 & b_1 - b_2 \\ 0 & -1 & -1 & b_3 - 2b_1 \end{bmatrix}$$

The second row was multiplied by -1.

$$\begin{bmatrix} 1 & 1 & 2 & b_1 \\ 0 & 1 & 1 & b_1 - b_2 \\ 0 & 0 & 0 & b_3 - b_2 - b_1 \end{bmatrix}$$

The second row was added to the third.

It is now evident from the third row in the matrix that the system has a solution if and only if b_1, b_2, and b_3 satisfy the condition

$$b_3 - b_2 - b_1 = 0 \quad \text{or} \quad b_3 = b_1 + b_2$$

To express this condition another way, $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is a matrix of the form

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_1 + b_2 \end{bmatrix}$$

where b_1 and b_2 are arbitrary.

EXAMPLE 4 | Determining Consistency by Elimination

What conditions must b_1, b_2, and b_3 satisfy in order for the system of equations

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= b_1 \\ 2x_1 + 5x_2 + 3x_3 &= b_2 \\ x_1 \qquad\quad + 8x_3 &= b_3 \end{aligned}$$

to be consistent?

Solution The augmented matrix is

$$\begin{bmatrix} 1 & 2 & 3 & b_1 \\ 2 & 5 & 3 & b_2 \\ 1 & 0 & 8 & b_3 \end{bmatrix}$$

Reducing this to reduced row echelon form yields (verify)

$$\begin{bmatrix} 1 & 0 & 0 & -40b_1 + 16b_2 + 9b_3 \\ 0 & 1 & 0 & 13b_1 - 5b_2 - 3b_3 \\ 0 & 0 & 1 & 5b_1 - 2b_2 - b_3 \end{bmatrix} \tag{2}$$

In this case there are no restrictions on b_1, b_2, and b_3, so the system has the unique solution

$$x_1 = -40b_1 + 16b_2 + 9b_3, \quad x_2 = 13b_1 - 5b_2 - 3b_3, \quad x_3 = 5b_1 - 2b_2 - b_3 \tag{3}$$

for all values of b_1, b_2, and b_3.

What does the result in Example 4 tell you about the coefficient matrix of the system?

Exercise Set 1.6

In Exercises 1–8, solve the system by inverting the coefficient matrix and using Theorem 1.6.2.

1. $\begin{aligned} x_1 + x_2 &= 2 \\ 5x_1 + 6x_2 &= 9 \end{aligned}$

2. $\begin{aligned} 4x_1 - 3x_2 &= -3 \\ 2x_1 - 5x_2 &= 9 \end{aligned}$

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

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

5. $\begin{aligned} x + y + z &= 5 \\ x + y - 4z &= 10 \\ -4x + y + z &= 0 \end{aligned}$

6. $\begin{aligned} -x - 2y - 3z &= 0 \\ w + x + 4y + 4z &= 7 \\ w + 3x + 7y + 9z &= 4 \\ -w - 2x - 4y - 6z &= 6 \end{aligned}$

7. $\begin{aligned} 3x_1 + 5x_2 &= b_1 \\ x_1 + 2x_2 &= b_2 \end{aligned}$

8. $\begin{aligned} x_1 + 2x_2 + 3x_3 &= b_1 \\ 2x_1 + 5x_2 + 5x_3 &= b_2 \\ 3x_1 + 5x_2 + 8x_3 &= b_3 \end{aligned}$

In Exercises 9–12, solve the linear systems. Using the given values for the b's solve the systems together by reducing an appropriate augmented matrix to reduced row echelon form.

9. $\begin{aligned} x_1 - 5x_2 &= b_1 \\ 3x_1 + 2x_2 &= b_2 \end{aligned}$

i. $b_1 = 1, \; b_2 = 4$ **ii.** $b_1 = -2, \; b_2 = 5$

10. $\begin{aligned} -x_1 + 4x_2 + x_3 &= b_1 \\ x_1 + 9x_2 - 2x_3 &= b_2 \\ 6x_1 + 4x_2 - 8x_3 &= b_3 \end{aligned}$

i. $b_1 = 0, \; b_2 = 1, \; b_3 = 0$ **ii.** $b_1 = -3, \; b_2 = 4, \; b_3 = -5$

11. $\begin{aligned} 4x_1 - 7x_2 &= b_1 \\ x_1 + 2x_2 &= b_2 \end{aligned}$

i. $b_1 = 0, \quad b_2 = 1$ **ii.** $b_1 = -4, \quad b_2 = 6$
iii. $b_1 = -1, \quad b_2 = 3$ **iv.** $b_1 = -5, \quad b_2 = 1$

12. $\begin{aligned} x_1 + 3x_2 + 5x_3 &= b_1 \\ -x_1 - 2x_2 &= b_2 \\ 2x_1 + 5x_2 + 4x_3 &= b_3 \end{aligned}$

i. $b_1 = 1, \quad b_2 = 0, \quad b_3 = -1$
ii. $b_1 = 0, \quad b_2 = 1, \quad b_3 = 1$
iii. $b_1 = -1, \quad b_2 = -1, \quad b_3 = 0$

In Exercises 13–17, determine conditions on the b_i's, if any, in order to guarantee that the linear system is consistent.

13. $\begin{aligned} x_1 + 3x_2 &= b_1 \\ -2x_1 + x_2 &= b_2 \end{aligned}$

14. $\begin{aligned} 6x_1 - 4x_2 &= b_1 \\ 3x_1 - 2x_2 &= b_2 \end{aligned}$

15. $\begin{aligned} x_1 - 2x_2 + 5x_3 &= b_1 \\ 4x_1 - 5x_2 + 8x_3 &= b_2 \\ -3x_1 + 3x_2 - 3x_3 &= b_3 \end{aligned}$

16. $\begin{aligned} x_1 - 2x_2 - x_3 &= b_1 \\ -4x_1 + 5x_2 + 2x_3 &= b_2 \\ -4x_1 + 7x_2 + 4x_3 &= b_3 \end{aligned}$

17. $\begin{aligned} x_1 - x_2 + 3x_3 + 2x_4 &= b_1 \\ -2x_1 + x_2 + 5x_3 + x_4 &= b_2 \\ -3x_1 + 2x_2 + 2x_3 - x_4 &= b_3 \\ 4x_1 - 3x_2 + x_3 + 3x_4 &= b_4 \end{aligned}$

18. Consider the matrices

$$A = \begin{bmatrix} 2 & 1 & 2 \\ 2 & 2 & -2 \\ 3 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

a. Show that the equation $A\mathbf{x} = \mathbf{x}$ can be rewritten as $(A - I)\mathbf{x} = \mathbf{0}$ and use this result to solve $A\mathbf{x} = \mathbf{x}$ for $\mathbf{x}$.

b. Solve $A\mathbf{x} = 4\mathbf{x}$.

In Exercises 19–20, solve the matrix equation for X.

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

20. $\begin{bmatrix} -2 & 0 & 1 \\ 0 & -1 & -1 \\ 1 & 1 & -4 \end{bmatrix} X = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 6 & 7 & 8 & 9 \\ 1 & 3 & 7 & 9 \end{bmatrix}$

Working with Proofs

21. Let $A\mathbf{x} = \mathbf{0}$ be a homogeneous system of n linear equations in n unknowns that has only the trivial solution. Prove that if k is any positive integer, then the system $A^k\mathbf{x} = \mathbf{0}$ also has only the trivial solution.

22. Let $A\mathbf{x} = \mathbf{0}$ be a homogeneous system of n linear equations in n unknowns, and let Q be an invertible $n \times n$ matrix. Prove that $A\mathbf{x} = \mathbf{0}$ has only the trivial solution if and only if $(QA)\mathbf{x} = \mathbf{0}$ has only the trivial solution.

23. Let $A\mathbf{x} = \mathbf{b}$ be any consistent system of linear equations, and let $\mathbf{x}_1$ be a fixed solution. Prove that every solution to the system can be written in the form $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_0$, where $\mathbf{x}_0$ is a solution to $A\mathbf{x} = \mathbf{0}$. Prove also that every matrix of this form is a solution.

24. Use part (a) of Theorem 1.6.3 to prove part (b).

True-False Exercises

TF. In parts (a)–(g) determine whether the statement is true or false, and justify your answer.

a. It is impossible for a system of linear equations to have exactly two solutions.

b. If A is a square matrix, and if the linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution, then the linear system $A\mathbf{x} = \mathbf{c}$ also must have a unique solution.

c. If A and B are $n \times n$ matrices such that $AB = I_n$, then $BA = I_n$.

d. If A and B are row equivalent matrices, then the linear systems $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ have the same solution set.

e. Let A be an $n \times n$ matrix and S is an $n \times n$ invertible matrix. If $\mathbf{x}$ is a solution to the system $(S^{-1}AS)\mathbf{x} = \mathbf{b}$, then $S\mathbf{x}$ is a solution to the system $A\mathbf{y} = S\mathbf{b}$.

f. Let A be an $n \times n$ matrix. The linear system $A\mathbf{x} = 4\mathbf{x}$ has a unique solution if and only if $A - 4I$ is an invertible matrix.

g. Let A and B be $n \times n$ matrices. If A or B (or both) are not invertible, then neither is AB.

Working with Technology

T1. Colors in print media, on computer monitors, and on television screens are implemented using what are called "color models." For example, in the RGB model, colors are created by mixing percentages of red (R), green (G), and blue (B), and in the YIQ model (used in TV broadcasting), colors are created by mixing percentages of luminescence (Y) with percentages of a chrominance factor (I) and a chrominance factor (Q). The conversion from the RGB model to the YIQ model is accomplished by the matrix equation

$$\begin{bmatrix} \text{Y} \\ \text{I} \\ \text{Q} \end{bmatrix} = \begin{bmatrix} .299 & .587 & .114 \\ .596 & -.275 & -.321 \\ .212 & -.523 & .311 \end{bmatrix} \begin{bmatrix} \text{R} \\ \text{G} \\ \text{B} \end{bmatrix}$$

What matrix would you use to convert the YIQ model to the RGB model?

T2. Let

$$A = \begin{bmatrix} 1 & -2 & 2 \\ 4 & 5 & 1 \\ 0 & 3 & -1 \end{bmatrix}, \; B_1 = \begin{bmatrix} 0 \\ 1 \\ 7 \end{bmatrix}, \; B_2 = \begin{bmatrix} 11 \\ 5 \\ 3 \end{bmatrix}, \; B_3 = \begin{bmatrix} 1 \\ -4 \\ 2 \end{bmatrix}$$

Solve the linear systems $A\mathbf{x} = B_1, A\mathbf{x} = B_2, A\mathbf{x} = B_3$ using the method of Example 2.

1.7 Diagonal, Triangular, and Symmetric Matrices

In this section we will discuss matrices that have various special forms. These matrices arise in a wide variety of applications and will play an important role in our subsequent work.

Diagonal Matrices

A square matrix in which all the entries off the main diagonal are zero is called a ***diagonal matrix***. Here are some examples:

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

A general $n \times n$ diagonal matrix D can be written as

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \tag{1}$$

A diagonal matrix is invertible if and only if all of its diagonal entries are nonzero; in this case the inverse of (1) is

$$D^{-1} = \begin{bmatrix} 1/d_1 & 0 & \cdots & 0 \\ 0 & 1/d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1/d_n \end{bmatrix} \tag{2}$$

We leave it for you to confirm that $DD^{-1} = D^{-1}D = I_m$.

Powers of diagonal matrices are easy to compute; we also leave it for you to verify that if D is the diagonal matrix (1) and k is a positive integer, then

$$D^k = \begin{bmatrix} d_1^k & 0 & \cdots & 0 \\ 0 & d_2^k & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_n^k \end{bmatrix} \tag{3}$$

EXAMPLE 1 | Inverses and Powers of Diagonal Matrices

If

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

then

$$A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}, \quad A^5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -243 & 0 \\ 0 & 0 & 32 \end{bmatrix}, \quad A^{-5} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{243} & 0 \\ 0 & 0 & \frac{1}{32} \end{bmatrix}$$

Matrix products that involve diagonal factors are especially easy to compute. For example,

$$\begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} = \begin{bmatrix} d_1a_{11} & d_1a_{12} & d_1a_{13} & d_1a_{14} \\ d_2a_{21} & d_2a_{22} & d_2a_{23} & d_2a_{24} \\ d_3a_{31} & d_3a_{32} & d_3a_{33} & d_3a_{34} \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}\begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} = \begin{bmatrix} d_1a_{11} & d_2a_{12} & d_3a_{13} \\ d_1a_{21} & d_2a_{22} & d_3a_{23} \\ d_1a_{31} & d_2a_{32} & d_3a_{33} \\ d_1a_{41} & d_2a_{42} & d_3a_{43} \end{bmatrix}$$

In words, *to multiply a matrix A on the left by a diagonal matrix D, multiply successive rows of A by the successive diagonal entries of D, and to multiply A on the right by D, multiply successive columns of A by the successive diagonal entries of D.*

Triangular Matrices

A square matrix in which all the entries above the main diagonal are zero is called ***lower triangular***, and a square matrix in which all the entries below the main diagonal are zero is called ***upper triangular***. A matrix that is either upper triangular or lower triangular is called ***triangular***.

EXAMPLE 2 | Upper and Lower Triangular Matrices

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix} \qquad \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

A general 4×4 upper triangular matrix

A general 4×4 lower triangular matrix

Remark Observe that diagonal matrices are both upper triangular and lower triangular since they have zeros below and above the main diagonal. Observe also that a *square* matrix in row echelon form is upper triangular since it has zeros below the main diagonal.

Properties of Triangular Matrices

Example 2 illustrates the following four facts about triangular matrices that we will state without formal proof:

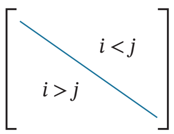


FIGURE 1.7.1

- A square matrix $A = [a_{ij}]$ is upper triangular if and only if all entries below the main diagonal are zero; that is, $a_{ij} = 0$ if $i > j$ (**Figure 1.7.1**).
- A square matrix $A = [a_{ij}]$ is lower triangular if and only if all entries above the main diagonal are zero; that is, $a_{ij} = 0$ if $i < j$ (Figure 1.7.1).
- A square matrix $A = [a_{ij}]$ is upper triangular if and only if the ith row starts with at least $i - 1$ zeros for every i.
- A square matrix $A = [a_{ij}]$ is lower triangular if and only if the jth column starts with at least $j - 1$ zeros for every j.

The following theorem lists some of the basic properties of triangular matrices.

Theorem 1.7.1

(*a*) The transpose of a lower triangular matrix is upper triangular, and the transpose of an upper triangular matrix is lower triangular.

(*b*) The product of lower triangular matrices is lower triangular, and the product of upper triangular matrices is upper triangular.

(*c*) A triangular matrix is invertible if and only if its diagonal entries are all nonzero.

(*d*) The inverse of an invertible lower triangular matrix is lower triangular, and the inverse of an invertible upper triangular matrix is upper triangular.

Part (*a*) is evident from the fact that transposing a square matrix can be accomplished by reflecting the entries about the main diagonal; we omit the formal proof. We will prove (*b*), but we will defer the proofs of (*c*) and (*d*) to the next chapter, where we will have the tools to prove those results more efficiently.

Proof (b) We will prove the result for lower triangular matrices; the proof for upper triangular matrices is similar. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be lower triangular $n \times n$ matrices, and let $C = [c_{ij}]$ be the product $C = AB$. We can prove that C is lower triangular by showing that $c_{ij} = 0$ for $i < j$. But from the definition of matrix multiplication,

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

If we assume that $i < j$, then the terms in this expression can be grouped as follows:

$$c_{ij} = \underbrace{a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{i(j-1)}b_{(j-1)j}}_{\substack{\text{Terms in which the row} \\ \text{number of } b \text{ is less than} \\ \text{the column number of } b}} + \underbrace{a_{ij}b_{jj} + \cdots + a_{in}b_{nj}}_{\substack{\text{Terms in which the row} \\ \text{number of } a \text{ is less than} \\ \text{the column number of } a}}$$

In the first grouping all of the b factors are zero since B is lower triangular, and in the second grouping all of the a factors are zero since A is lower triangular. Thus, $c_{ij} = 0$, which is what we wanted to prove. ■

EXAMPLE 3 | Computations with Triangular Matrices

Consider the upper triangular matrices

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

It follows from part (*c*) of Theorem 1.7.1 that the matrix A is invertible but the matrix B is not. Moreover, the theorem also tells us that A^{-1}, AB, and BA must be upper triangular. We leave it for you to confirm these three statements by showing that

$$A^{-1} = \begin{bmatrix} 1 & -\frac{3}{2} & \frac{7}{5} \\ 0 & \frac{1}{2} & -\frac{2}{5} \\ 0 & 0 & \frac{1}{5} \end{bmatrix}, \quad AB = \begin{bmatrix} 3 & -2 & -2 \\ 0 & 0 & 2 \\ 0 & 0 & 5 \end{bmatrix}, \quad BA = \begin{bmatrix} 3 & 5 & -1 \\ 0 & 0 & -5 \\ 0 & 0 & 5 \end{bmatrix}$$

Remark Observe that in this example the diagonal entries of AB and BA are the same and are the products of the corresponding diagonal entries of A and B. Also observe that the diagonal entries of A^{-1} are the reciprocals of the diagonal entries of A. In the exercises we ask you to show that this happens whenever upper or lower triangular matrices are multiplied or inverted.

Symmetric Matrices

Definition 1

A square matrix A is said to be ***symmetric*** if $A = A^T$.

It is easy to recognize a symmetric matrix by inspection: The entries on the main diagonal have no restrictions, but mirror images of entries *across* the main diagonal must be equal. Here is a picture using the second matrix in Example 4:

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

EXAMPLE 4 | Symmetric Matrices

The following matrices are symmetric since each is equal to its own transpose (verify).

$$\begin{bmatrix} 7 & -3 \\ -3 & 5 \end{bmatrix}, \quad \begin{bmatrix} 1 & 4 & 5 \\ 4 & -3 & 0 \\ 5 & 0 & 7 \end{bmatrix}, \quad \begin{bmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & d_3 & 0 \\ 0 & 0 & 0 & d_4 \end{bmatrix}$$

Remark It follows from Formula (14) of Section 1.3 that a square matrix A is symmetric if and only if

$$\boxed{(A)_{ij} = (A)_{ji}} \tag{4}$$

for all values of i and j.

The following theorem lists the main algebraic properties of symmetric matrices. The proofs are direct consequences of Theorem 1.4.8 and are omitted.

Theorem 1.7.2

If A and B are symmetric matrices with the same size, and if k is any scalar, then:

(*a*) A^T is symmetric.

(*b*) $A + B$ and $A - B$ are symmetric.

(*c*) kA is symmetric.

It is not true, in general, that the product of symmetric matrices is symmetric. To see why this is so, let A and B be symmetric matrices with the same size. Then it follows from part (*e*) of Theorem 1.4.8 and the symmetry of A and B that

$$(AB)^T = B^T A^T = BA$$

Thus, $(AB)^T = AB$ if and only if $AB = BA$, that is, if and only if A and B commute. In summary, we have the following result.

Theorem 1.7.3

The product of two symmetric matrices is symmetric if and only if the matrices commute.

EXAMPLE 5 | Products of Symmetric Matrices

The first of the following equations shows a product of symmetric matrices that *is not* symmetric, and the second shows a product of symmetric matrices that *is* symmetric. We conclude that the factors in the first equation do not commute, but those in the second equation do. We leave it for you to verify that this is so.

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

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

Invertibility of Symmetric Matrices

In general, a symmetric matrix need not be invertible. For example, a diagonal matrix with a zero on the main diagonal is symmetric but not invertible. However, the following theorem shows that if a symmetric matrix happens to be invertible, then its inverse must also be symmetric.

Theorem 1.7.4

If A is an invertible symmetric matrix, then A^{-1} is symmetric.

Proof Assume that A is symmetric and invertible. From Theorem 1.4.9 and the fact that $A = A^T$, we have

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

which proves that A^{-1} is symmetric. ■

Later in this text, we will obtain general conditions on A under which AA^T and A^TA are invertible. However, in the special case where A is *square*, we have the following result.

Theorem 1.7.5

If A is an invertible matrix, then AA^T and A^TA are also invertible.

Proof Since A is invertible, so is A^T by Theorem 1.4.9. Thus AA^T and A^TA are invertible, since they are the products of invertible matrices. ■

Products AA^T and A^TA are Symmetric

Matrix products of the form AA^T and A^TA arise in a variety of applications. If A is an $m \times n$ matrix, then A^T is an $n \times m$ matrix, so the products AA^T and A^TA are both square matrices—the matrix AA^T has size $m \times m$, and the matrix A^TA has size $n \times n$. Such products are always symmetric since

$$(AA^T)^T = (A^T)^TA^T = AA^T \quad \text{and} \quad (A^TA)^T = A^T(A^T)^T = A^TA$$

EXAMPLE 6 | **The Product of a Matrix and Its Transpose Is Symmetric**

Let A be the 2×3 matrix

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

Then

$$A^TA = \begin{bmatrix} 1 & 3 \\ -2 & 0 \\ 4 & -5 \end{bmatrix} \begin{bmatrix} 1 & -2 & 4 \\ 3 & 0 & -5 \end{bmatrix} = \begin{bmatrix} 10 & -2 & -11 \\ -2 & 4 & -8 \\ -11 & -8 & 41 \end{bmatrix}$$

$$AA^T = \begin{bmatrix} 1 & -2 & 4 \\ 3 & 0 & -5 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -2 & 0 \\ 4 & -5 \end{bmatrix} = \begin{bmatrix} 21 & -17 \\ -17 & 34 \end{bmatrix}$$

Observe that A^TA and AA^T are symmetric as expected.

Exercise Set 1.7

In Exercises 1–2, classify the matrix as upper triangular, lower triangular, or diagonal, and decide by inspection whether the matrix is invertible. Recall that a diagonal matrix is both upper and lower triangular, so there may be more than one answer in some parts.

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

c. $\begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix}$ **d.** $\begin{bmatrix} 3 & -2 & 7 \\ 0 & 0 & 3 \\ 0 & 0 & 8 \end{bmatrix}$

2. a. $\begin{bmatrix} 4 & 0 \\ 1 & 7 \end{bmatrix}$ **b.** $\begin{bmatrix} 0 & -3 \\ 0 & 0 \end{bmatrix}$

c. $\begin{bmatrix} 4 & 0 & 0 \\ 0 & \frac{3}{5} & 0 \\ 0 & 0 & -2 \end{bmatrix}$ **d.** $\begin{bmatrix} 3 & 0 & 0 \\ 3 & 1 & 0 \\ 7 & 0 & 0 \end{bmatrix}$

In Exercises 3–6, find the product by inspection.

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

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

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

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

In Exercises 7–10, find A^2, A^{-2}, and A^{-k} (where k is any integer) by inspection.

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

9. $A = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix}$ **10.** $A = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & -4 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$

In Exercises 11–12, compute the product by inspection.

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

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

In Exercises 13–14, compute the indicated quantity.

13. $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}^{39}$ **14.** $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}^{1000}$

In Exercises 15–16, use what you have learned in this section about multiplying by diagonal matrices to compute the product by inspection.

15. a. $\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}\begin{bmatrix} u & v \\ w & x \\ y & z \end{bmatrix}$ **b.** $\begin{bmatrix} r & s & t \\ u & v & w \\ x & y & z \end{bmatrix}\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$

16. a. $\begin{bmatrix} u & v \\ w & x \\ y & z \end{bmatrix}\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$ **b.** $\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}\begin{bmatrix} r & s & t \\ u & v & w \\ x & y & z \end{bmatrix}$

In Exercises 17–18, create a symmetric matrix by substituting appropriate numbers for the ×'s.

17. a. $\begin{bmatrix} 2 & -1 \\ \times & 3 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & \times & \times & \times \\ 3 & 1 & \times & \times \\ 7 & -8 & 0 & \times \\ 2 & -3 & 9 & 0 \end{bmatrix}$

18. a. $\begin{bmatrix} 0 & \times \\ 3 & 0 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 7 & -3 & 2 \\ \times & 4 & 5 & -7 \\ \times & \times & 1 & -6 \\ \times & \times & \times & 3 \end{bmatrix}$

In Exercises 19–22, determine by inspection whether the matrix is invertible.

19. $\begin{bmatrix} 0 & 6 & -1 \\ 0 & 7 & -4 \\ 0 & 0 & -2 \end{bmatrix}$ **20.** $\begin{bmatrix} -1 & 2 & 4 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

21. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & -5 & 0 & 0 \\ 4 & -3 & 4 & 0 \\ 1 & -2 & 1 & 3 \end{bmatrix}$ **22.** $\begin{bmatrix} 2 & 0 & 0 & 0 \\ -3 & -1 & 0 & 0 \\ -4 & -6 & 0 & 0 \\ 0 & 3 & 8 & -5 \end{bmatrix}$

In Exercises 23–24, find the diagonal entries of AB by inspection.

23. $A = \begin{bmatrix} 3 & 2 & 6 \\ 0 & 1 & -2 \\ 0 & 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 2 & 7 \\ 0 & 5 & 3 \\ 0 & 0 & 6 \end{bmatrix}$

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

In Exercises 25–26, find all values of the unknown constant(s) for which A is symmetric.

25. $A = \begin{bmatrix} 4 & -3 \\ a+5 & -1 \end{bmatrix}$

26. $A = \begin{bmatrix} 2 & a-2b+2c & 2a+b+c \\ 3 & 5 & a+c \\ 0 & -2 & 7 \end{bmatrix}$

In Exercises 27–28, find all values of x for which A is invertible.

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

28. $A = \begin{bmatrix} x-\frac{1}{2} & 0 & 0 \\ x & x-\frac{1}{3} & 0 \\ x^2 & x^3 & x+\frac{1}{4} \end{bmatrix}$

29. If A is an invertible upper triangular or lower triangular matrix, what can you say about the diagonal entries of A^{-1}?

30. Show that if A is a symmetric $n \times n$ matrix and B is any $n \times m$ matrix, then the following products are symmetric:

$$B^TB, \quad BB^T, \quad B^TAB$$

In Exercises 31–32, find a diagonal matrix A that satisfies the given condition.

31. $A^5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ **32.** $A^{-2} = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

33. Verify Theorem 1.7.1(b) for the matrix product AB and Theorem 1.7.1(d) for the matrix A, where

$$A = \begin{bmatrix} -1 & 2 & 5 \\ 0 & 1 & 3 \\ 0 & 0 & -4 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -8 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$$

34. Let A be an $n \times n$ symmetric matrix.

a. Show that A^2 is symmetric.

b. Show that $2A^2 - 3A + I$ is symmetric.

35. Verify Theorem 1.7.4 for the given matrix A.

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

36. Find all 3×3 diagonal matrices A that satisfy $A^2 - 3A - 4I = 0$.

37. Let $A = [a_{ij}]$ be an $n \times n$ matrix. Determine whether A is symmetric.

a. $a_{ij} = i^2 + j^2$ **b.** $a_{ij} = i^2 - j^2$

c. $a_{ij} = 2i + 2j$ **d.** $a_{ij} = 2i^2 + 2j^3$

38. On the basis of your experience with Exercise 37, devise a general test that can be applied to a formula for a_{ij} to determine whether $A = [a_{ij}]$ is symmetric.

39. Find an upper triangular matrix that satisfies

$$A^3 = \begin{bmatrix} 1 & 30 \\ 0 & -8 \end{bmatrix}$$

40. If the $n \times n$ matrix A can be expressed as $A = LU$, where L is a lower triangular matrix and U is an upper triangular matrix, then the linear system $A\mathbf{x} = \mathbf{b}$ can be expressed as $LU\mathbf{x} = \mathbf{b}$ and can be solved in two steps:

Step 1. Let $U\mathbf{x} = \mathbf{y}$, so that $LU\mathbf{x} = \mathbf{b}$ can be expressed as $L\mathbf{y} = \mathbf{b}$. Solve this system.

Step 2. Solve the system $U\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$.

In each part, use this two-step method to solve the given system.

a. $\begin{bmatrix} 1 & 0 & 0 \\ -2 & 3 & 0 \\ 2 & 4 & 1 \end{bmatrix}\begin{bmatrix} 2 & -1 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}$

b. $\begin{bmatrix} 2 & 0 & 0 \\ 4 & 1 & 0 \\ -3 & -2 & 3 \end{bmatrix}\begin{bmatrix} 3 & -5 & 2 \\ 0 & 4 & 1 \\ 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ -5 \\ 2 \end{bmatrix}$

In the text we defined a matrix A to be symmetric if $A^T = A$. Analogously, a matrix A is said to be **skew-symmetric** *if $A^T = -A$. Exercises 41–45 are concerned with matrices of this type.*

41. Fill in the missing entries (marked with $\times$) so the matrix A is skew-symmetric.

a. $A = \begin{bmatrix} \times & \times & 4 \\ 0 & \times & \times \\ \times & -1 & \times \end{bmatrix}$ **b.** $A = \begin{bmatrix} \times & 0 & \times \\ \times & \times & -4 \\ 8 & \times & \times \end{bmatrix}$

42. Find all values of a, b, c, and d for which A is skew-symmetric.

$$A = \begin{bmatrix} 0 & 2a - 3b + c & 3a - 5b + 5c \\ -2 & 0 & 5a - 8b + 6c \\ -3 & -5 & d \end{bmatrix}$$

43. We showed in the text that the product of symmetric matrices is symmetric if and only if the matrices commute. Is the product of commuting skew-symmetric matrices skew-symmetric? Explain.

Working with Proofs

44. Prove that every square matrix A can be expressed as the sum of a symmetric matrix and a skew-symmetric matrix. [*Hint:* Note the identity $A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$.]

45. Prove the following facts about skew-symmetric matrices.

a. If A is an invertible skew-symmetric matrix, then A^{-1} is skew-symmetric.

b. If A and B are skew-symmetric matrices, then so are A^T, $A + B$, $A - B$, and kA for any scalar k.

46. Prove: If the matrices A and B are both upper triangular or both lower triangular, then the diagonal entries of both AB and BA are the products of the diagonal entries of A and B.

47. Prove: If $A^TA = A$, then A is symmetric and $A = A^2$.

True-False Exercises

TF. In parts (a)–(m) determine whether the statement is true or false, and justify your answer.

a. The transpose of a diagonal matrix is a diagonal matrix.

b. The transpose of an upper triangular matrix is an upper triangular matrix.

c. The sum of an upper triangular matrix and a lower triangular matrix is a diagonal matrix.

d. All entries of a symmetric matrix are determined by the entries occurring on and above the main diagonal.

e. All entries of an upper triangular matrix are determined by the entries occurring on and above the main diagonal.

f. The inverse of an invertible lower triangular matrix is an upper triangular matrix.

g. A diagonal matrix is invertible if and only if all of its diagonal entries are positive.

h. The sum of a diagonal matrix and a lower triangular matrix is a lower triangular matrix.

i. A matrix that is both symmetric and upper triangular must be a diagonal matrix.

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

k. If A and B are $n \times n$ matrices such that $A + B$ is upper triangular, then A and B are upper triangular.

l. If A^2 is a symmetric matrix, then A is a symmetric matrix.

m. If kA is a symmetric matrix for some $k \neq 0$, then A is a symmetric matrix.

Working with Technology

T1. Starting with the formula stated in Exercise T1 of Section 1.5, derive a formula for the inverse of the "block diagonal" matrix

$$\begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}$$

in which D_1 and D_2 are invertible, and use your result to compute the inverse of the matrix

$$M = \begin{bmatrix} 1.24 & 2.37 & 0 & 0 \\ 3.08 & -1.01 & 0 & 0 \\ 0 & 0 & 2.76 & 4.92 \\ 0 & 0 & 3.23 & 5.54 \end{bmatrix}$$

1.8 Introduction to Linear Transformations

Up to now we have treated matrices simply as rectangular arrays of numbers and have been concerned primarily with developing algebraic properties of those arrays. In this section we will view matrices in a completely different way. Here we will be concerned with how matrices can be used to transform or "map" one vector into another by matrix multiplication. This will be the foundation for much of our work in subsequent sections.

Recall that in Section 1.1 we defined an "ordered n-tuple" to be a sequence of n real numbers, and we observed that a solution of a linear system in n unknowns, say

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

can be expressed as the ordered n-tuple

$$(s_1, s_2, \dots, s_n) \tag{1}$$

Recall also that if $n = 2$, then the n-tuple is called an "ordered pair," and if $n = 3$, it is called an "ordered triple." For two ordered n-tuples to be regarded as the same, they must list the same numbers in the same order. Thus, for example, $(1, 2)$ and $(2, 1)$ are different ordered pairs.

The set of all ordered n-tuples of real numbers is denoted by the symbol R^n. The elements of R^n are called ***vectors*** and are denoted in boldface type, such as $\mathbf{a}$, $\mathbf{b}$, $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{x}$. When convenient, ordered n-tuples can be denoted in matrix notation as column vectors. For example, the matrix

The term "vector" is used in various ways in mathematics, physics, engineering, and other applications. The idea of viewing n-tuples as vectors will be discussed in more detail in Chapter 3, at which point we will also explain how this idea relates to a more familiar notion of a vector.

$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} \tag{2}$$

can be used as an alternative to (1). We call (1) the ***comma-delimited form*** of a vector and (2) the ***column-vector form***. For each $i = 1, 2, \dots, n$, let $\mathbf{e}_i$ denote the vector in R^n with a 1 in the ith position and zeros elsewhere. In column form these vectors are

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

We call the vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ the ***standard basis vectors*** for R^n. For example, the vectors

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

are the standard basis vectors for R^3.

The vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ in R^n are termed "basis vectors" because all other vectors in R^n are expressible in exactly one way as a linear combination of them. For example, if

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

then we can express $\mathbf{x}$ as

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$$

Functions and Transformations

Recall that a ***function*** is a rule that associates with each element of a set A one and only one element in a set B. If f associates the element b with the element a, then we write

$$b = f(a)$$

and we say that b is the ***image*** of a under f or that $f(a)$ is the ***value*** of f at a. The set A is called the ***domain*** of f and the set B the ***codomain*** of f (**Figure 1.8.1**). The subset of the codomain that consists of all images of elements in the domain is called the ***range*** of f.

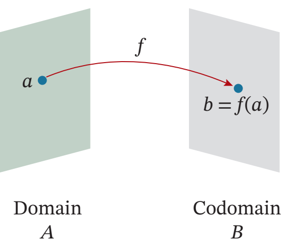


FIGURE 1.8.1

In many applications the domain and codomain of a function are sets of real numbers, but in this text we will be concerned with functions for which the domain is R^n and the codomain is R^m for some positive integers m and n. In this setting it is common to use italicized capital letters for functions, the letter T being typical.

Definition 1

If T is a function with domain R^n and codomain R^m, then we say that T is a ***transformation*** from R^n to R^m or that T ***maps*** from R^n to R^m, which we denote by writing

$$T : R^n \to R^m$$

In the special case where $m = n$, a transformation is sometimes called an ***operator*** on R^n.

Matrix Transformations

In this section we will be concerned with the class of transformations from R^n to R^m that arise from linear systems. Specifically, suppose that we have the system of linear equations

$$\begin{aligned}
w_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\
w_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\
&\;\;\vdots \\
w_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n
\end{aligned} \tag{3}$$

which we can write in matrix notation as

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{4}$$

or more briefly as

$$\mathbf{w} = A\mathbf{x} \tag{5}$$

Up to now we have been viewing (5) as a compact way of writing system (3). Another way to view this formula is as a transformation that maps a vector $\mathbf{x}$ in R^n into a vector $\mathbf{w}$ in R^m by multiplying $\mathbf{x}$ on the left by A. We call this a ***matrix transformation*** (or ***matrix operator*** in the special case where $m = n$). We denote it by

$$T_A : R^n \to R^m$$

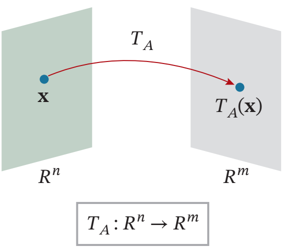


FIGURE 1.8.2

(see **Figure 1.8.2**). This notation is useful when it is important to make the domain and codomain clear. The subscript on T_A serves as a reminder that the transformation results from multiplying vectors in R^n by the matrix A. In situations where specifying the domain and codomain is not essential, we will express (5) as

$$\mathbf{w} = T_A(\mathbf{x}) \tag{6}$$

We call the transformation T_A ***multiplication by A***. On occasion we will find it convenient to express (6) in the schematic form

$$\mathbf{x} \xrightarrow{T_A} \mathbf{w} \tag{7}$$

which is read "T_A maps $\mathbf{x}$ into $\mathbf{w}$."

EXAMPLE 1 | A Matrix Transformation from R^4 to R^3

The transformation from R^4 to R^3 defined by the equations

$$\begin{aligned}
w_1 &= 2x_1 - 3x_2 + x_3 - 5x_4 \\
w_2 &= 4x_1 + x_2 - 2x_3 + x_4 \\
w_3 &= 5x_1 - x_2 + 4x_3
\end{aligned} \tag{8}$$

can be expressed in matrix form as

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 2 & -3 & 1 & -5 \\ 4 & 1 & -2 & 1 \\ 5 & -1 & 4 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

from which we see that the transformation can be interpreted as multiplication by

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

Although the image under the transformation T_A of any vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

in R^4 could be computed directly from the defining equations in (8), we will find it preferable to use the matrix in (9). For example, if

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

then it follows from (9) that

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

EXAMPLE 2 | Zero Transformations

If 0 is the $m \times n$ zero matrix, then

$$T_0(\mathbf{x}) = 0\mathbf{x} = \mathbf{0}$$

so multiplication by zero maps every vector in R^n into the zero vector in R^m. We call T_0 the ***zero transformation*** from R^n to R^m.

EXAMPLE 3 | Identity Operators

If I is the $n \times n$ identity matrix, then

$$T_I(\mathbf{x}) = I\mathbf{x} = \mathbf{x}$$

so multiplication by I maps every vector in R^n to itself. We call T_I the ***identity operator*** on R^n.

Properties of Matrix Transformations

The following theorem lists four basic properties of matrix transformations that follow from properties of matrix multiplication.

Theorem 1.8.1

For every matrix A the matrix transformation $T_A : R^n \to R^m$ has the following properties for all vectors $\mathbf{u}$ and $\mathbf{v}$ and for every scalar k:

(*a*) $T_A(\mathbf{0}) = \mathbf{0}$

(*b*) $T_A(k\mathbf{u}) = kT_A(\mathbf{u})$ **[Homogeneity property]**

(*c*) $T_A(\mathbf{u}+\mathbf{v}) = T_A(\mathbf{u}) + T_A(\mathbf{v})$ **[Additivity property]**

(*d*) $T_A(\mathbf{u}-\mathbf{v}) = T_A(\mathbf{u}) - T_A(\mathbf{v})$

Proof All four parts are restatements from the transformation viewpoint of the following properties of matrix arithmetic given in Theorem 1.4.1:

$$A\mathbf{0} = \mathbf{0}, \quad A(k\mathbf{u}) = k(A\mathbf{u}), \quad A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v}, \quad A(\mathbf{u}-\mathbf{v}) = A\mathbf{u} - A\mathbf{v} \quad \blacksquare$$

It follows from parts (*b*) and (*c*) of Theorem 1.8.1 that a matrix transformation maps a linear combination of vectors in R^n into the corresponding linear combination of vectors in R^m in the sense that

$$\boxed{T_A(k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \cdots + k_r\mathbf{u}_r) = k_1T_A(\mathbf{u}_1) + k_2T_A(\mathbf{u}_2) + \cdots + k_rT_A(\mathbf{u}_r)} \tag{10}$$

Matrix transformations are not the only kinds of transformations. For example, if

$$\begin{aligned} w_1 &= x_1^2 + x_2^2 \\ w_2 &= x_1x_2 \end{aligned} \tag{11}$$

then there are no constants a, b, c, and d for which

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1^2 + x_2^2 \\ x_1x_2 \end{bmatrix}$$

so that the equations in (11) do not define a matrix transformation from R^2 to R^2.

This leads us to the following two questions.

Question 1. Are there algebraic properties of a transformation $T : R^n \to R^m$ that can be used to determine whether T is a matrix transformation?

Question 2. If we discover that a transformation $T : R^n \to R^m$ is a matrix transformation, how can we find a matrix A for which $T = T_A$?

The following theorem and its proof will provide the answers.

Theorem 1.8.2

$T : R^n \to R^m$ is a matrix transformation if and only if the following relationships hold for all vectors $\mathbf{u}$ and $\mathbf{v}$ in R^n and for every scalar k:

(i) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ **[Additivity property]**

(ii) $T(k\mathbf{u}) = kT(\mathbf{u})$ **[Homogeneity property]**

Proof If T is a matrix transformation, then properties (i) and (ii) follow respectively from parts (c) and (b) of Theorem 1.8.1.

Conversely, assume that properties (i) and (ii) hold. We must show that there exists an $m \times n$ matrix A such that

$$T(\mathbf{x}) = A\mathbf{x}$$

for every vector $\mathbf{x}$ in R^n. Recall that the derivation of Formula (10) used only the additivity and homogeneity properties of T_A. Since we are assuming that T has those properties, it must be true that

$$T(k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \cdots + k_r\mathbf{u}_r) = k_1T(\mathbf{u}_1) + k_2T(\mathbf{u}_2) + \cdots + k_rT(\mathbf{u}_r) \qquad (12)$$

for all scalars $k_1, k_2, \ldots, k_r$ and all vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$ in R^n. Let A be the matrix

$$A = [T(\mathbf{e}_1) \mid T(\mathbf{e}_2) \mid \cdots \mid T(\mathbf{e}_n)] \qquad (13)$$

where $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ are the standard basis vectors for R^n. It follows from Theorem 1.3.1 that $A\mathbf{x}$ is a linear combination of the columns of A in which the successive coefficients are the entries $x_1, x_2, \ldots, x_n$ of $\mathbf{x}$. That is,

$$A\mathbf{x} = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \cdots + x_nT(\mathbf{e}_n)$$

Using Formula (10) we can rewrite this as

$$A\mathbf{x} = T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n) = T(\mathbf{x})$$

which completes the proof. ■

The two properties listed in Theorem 1.8.2 are called ***linearity conditions***, and a transformation that satisfies these conditions is called a ***linear transformation***. Using this terminology Theorem 1.8.2 can be restated as follows.

Theorem 1.8.3

Every linear transformation from R^n to R^m is a matrix transformation and conversely every matrix transformation from R^n to R^m is a linear transformation.

Briefly stated, this theorem tells us that for transformations from R^n to R^m the terms "linear transformation" and "matrix transformation" are synonymous.

Depending on whether n-tuples and m-tuples are regarded as vectors or points, the geometric effect of a matrix transformation $T_A: R^n \to R^m$ is to map each vector (point) in R^n into a vector (point) in R^m (**Figure 1.8.3**).

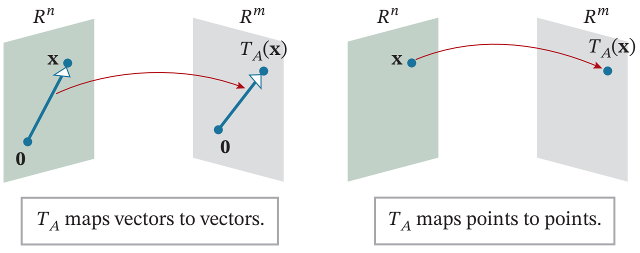


FIGURE 1.8.3

The following theorem states that if two matrix transformations from R^n to R^m have the same image for each point of R^n, then the matrices themselves must be the same.

Theorem 1.8.4

If $T_A: R^n \to R^m$ and $T_B: R^n \to R^m$ are matrix transformations, and if $T_A(\mathbf{x}) = T_B(\mathbf{x})$ for every vector $\mathbf{x}$ in R^n, then $A = B$.

Proof To say that $T_A(\mathbf{x}) = T_B(\mathbf{x})$ for every vector in R^n is the same as saying that

$$A\mathbf{x} = B\mathbf{x}$$

for every vector $\mathbf{x}$ in R^n. This will be true, in particular, if $\mathbf{x}$ is any of the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ for R^n; that is,

$$A\mathbf{e}_j = B\mathbf{e}_j \quad (j = 1, 2, \dots, n) \tag{14}$$

Since every entry of $\mathbf{e}_j$ is 0 except for the jth, which is 1, it follows from Theorem 1.3.1 that $A\mathbf{e}_j$ is the jth column of A and $B\mathbf{e}_j$ is the jth column of B. Thus, (14) implies that corresponding columns of A and B are the same, and hence that $A = B$. ■

Theorem 1.8.4 is significant because it tells us that there is a *one-to-one correspondence* between $m \times n$ matrices and matrix transformations from R^n to R^m in the sense that every $m \times n$ matrix A produces exactly one matrix transformation (multiplication by A) and every matrix transformation from R^n to R^m arises from exactly one $m \times n$ matrix; we call that matrix the ***standard matrix*** for the transformation.

A Procedure for Finding Standard Matrices

In the course of proving Theorem 1.8.2 we showed in Formula (13) that if $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ are the standard basis vectors for R^n (in column form), then the standard matrix for a linear transformation $T: R^n \to R^m$ is given by the formula

$$\boxed{A = [T(\mathbf{e}_1) \mid T(\mathbf{e}_2) \mid \cdots \mid T(\mathbf{e}_n)]} \tag{15}$$

This formula reveals a key property of linear transformations from R^n to R^m, namely, that they are completely determined by their actions on the standard basis vectors for R^n. It also suggests the following procedure that can be used to find the standard matrix for such transformations.

Finding the Standard Matrix for a Matrix Transformation

Step 1. Find the images of the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ for R^n.

Step 2. Construct the matrix that has the images obtained in Step 1 as its successive columns. This matrix is the standard matrix for the transformation.

EXAMPLE 4 | Finding a Standard Matrix

Find the standard matrix A for the linear transformation $T: R^2 \to R^3$ defined by the formula

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + x_2 \\ x_1 - 3x_2 \\ -x_1 + x_2 \end{bmatrix} \tag{16}$$

Solution We leave it for you to verify that

$$T(\mathbf{e}_1) = T\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}2\\1\\-1\end{bmatrix} \quad \text{and} \quad T(\mathbf{e}_2) = T\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{bmatrix}1\\-3\\1\end{bmatrix}$$

Thus, it follows from Formulas (15) and (16) that the standard matrix is

$$A = [T(\mathbf{e}_1) \mid T(\mathbf{e}_2)] = \begin{bmatrix}2 & 1\\1 & -3\\-1 & 1\end{bmatrix}$$

As a check, observe that

$$A\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}2 & 1\\1 & -3\\-1 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}2x_1 + x_2\\x_1 - 3x_2\\-x_1 + x_2\end{bmatrix}$$

which shows that multiplication by A produces the same result as the transformation T (see Equation (16)).

EXAMPLE 5 | Computing with Standard Matrices

For the linear transformation in Example 4, use the standard matrix A obtained in that example to find

$$T\left(\begin{bmatrix}1\\4\end{bmatrix}\right)$$

Solution The transformation is multiplication by A, so

$$T\left(\begin{bmatrix}1\\4\end{bmatrix}\right) = \begin{bmatrix}2 & 1\\1 & -3\\-1 & 1\end{bmatrix}\begin{bmatrix}1\\4\end{bmatrix} = \begin{bmatrix}6\\-11\\3\end{bmatrix}$$

Although we could have obtained the result in Example 5 by substituting values for the variables in (13), the method used in that example is preferable for large-scale problems in that matrix multiplication is better suited for computer computations.

For transformation problems posed in comma-delimited form, a good procedure is to rewrite the problem in column-vector form and use the methods previously illustrated.

EXAMPLE 6 | Finding a Standard Matrix

Rewrite the transformation $T(x_1, x_2) = (3x_1 + x_2, 2x_1 - 4x_2)$ in column-vector form and find its standard matrix.

Solution

$$T\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}3x_1 + x_2\\2x_1 - 4x_2\end{bmatrix} = \begin{bmatrix}3 & 1\\2 & -4\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}$$

Thus, the standard matrix is

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

EXAMPLE 7

Find the standard matrix A for the linear transformation $T: R^2 \to R^2$ for which

$$T\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -5 \\ 5 \end{bmatrix}, \quad T\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 7 \\ -6 \end{bmatrix} \tag{17}$$

Solution Our objective is to find the images of the standard basis vectors and then use Formula (15) to obtain the standard matrix. To start, we will rewrite the standard basis vectors as linear combinations of

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

This leads to the vector equations

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_1\begin{bmatrix} -1 \\ 1 \end{bmatrix} + c_2\begin{bmatrix} 2 \\ -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} = k_1\begin{bmatrix} -1 \\ 1 \end{bmatrix} + k_2\begin{bmatrix} 2 \\ -1 \end{bmatrix} \tag{18}$$

which we can rewrite as

$$\begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

As these systems have the same coefficient matrix, we can solve both at once using the method in Example 2 of Section 1.6. We leave it for you to do this and to show that

$$c_1 = 1, \; c_2 = 1, \; k_1 = 2, \; k_2 = 1$$

Substituting these values in (18) and using the linearity properties of T, we obtain

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = T\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) + T\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} -5 \\ 5 \end{bmatrix} + \begin{bmatrix} 7 \\ -6 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

$$T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = 2T\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) + T\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} -10 \\ 10 \end{bmatrix} + \begin{bmatrix} 7 \\ -6 \end{bmatrix} = \begin{bmatrix} -3 \\ 4 \end{bmatrix}$$

Thus, it follows from Formula (15) that the standard matrix for T is

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

You can check this result using multiplication by A to verify (17).

Remark This section is but a first step in the study of linear transformations, which is one of the major themes in this text. We will delve deeper into this topic in Chapter 4, at which point we will have more background and a richer source of examples to work with.

There are many ways to transform the vector spaces R^2 and R^3, some of the most important of which can be accomplished by matrix transformations. For example, rotations about the origin, reflections about lines and planes through the origin, and projections onto lines and planes through the origin can all be accomplished using a matrix operator with an appropriate 2×2 or 3×3 matrix.

Reflection Operators

Some of the most basic matrix operators on R^2 and R^3 are those that map each point into its symmetric image about a fixed line or a fixed plane that contains the origin; these are called ***reflection operators***. **Table 1** shows the standard matrices for the reflections about the coordinate axes and the line $y = x$ in R^2, and **Table 2** shows the standard matrices for the reflections about the coordinate planes in R^3. In each case the standard matrix was obtained by finding the images of the standard basis vectors, converting those images to column vectors, and then using those column vectors as successive columns of the standard matrix.

TABLE 1

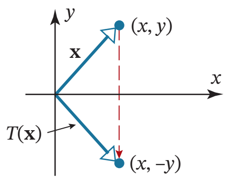
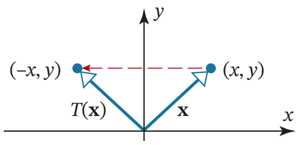
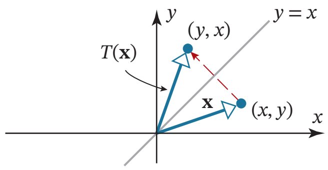
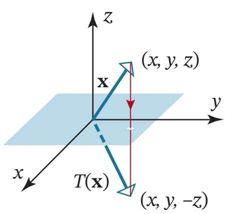
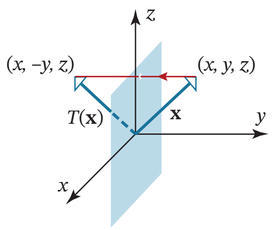
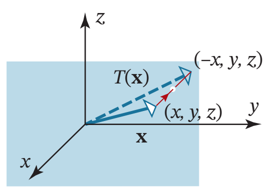
Operator	Illustration	Images of $\mathbf{e}_1$ and $\mathbf{e}_2$	Standard Matrix
Reflection about the x-axis $T(x, y) = (x, -y)$		$T(\mathbf{e}_1) = T(1, 0) = (1, 0)$ $T(\mathbf{e}_2) = T(0, 1) = (0, -1)$	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
Reflection about the y-axis $T(x, y) = (-x, y)$		$T(\mathbf{e}_1) = T(1, 0) = (-1, 0)$ $T(\mathbf{e}_2) = T(0, 1) = (0, 1)$	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$
Reflection about the line $y = x$ $T(x, y) = (y, x)$		$T(\mathbf{e}_1) = T(1, 0) = (0, 1)$ $T(\mathbf{e}_2) = T(0, 1) = (1, 0)$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

TABLE 2

Operator	Illustration	Images of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$	Standard Matrix
Reflection about the xy-plane $T(x, y, z) = (x, y, -z)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (1, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, 1, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, -1)$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$
Reflection about the xz-plane $T(x, y, z) = (x, -y, z)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (1, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, -1, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, 1)$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Reflection about the yz-plane $T(x, y, z) = (-x, y, z)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (-1, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, 1, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, 1)$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Projection Operators

Matrix operators on R^2 and R^3 that map each point into its orthogonal projection onto a fixed line or plane through the origin are called ***projection operators*** (or more precisely, ***orthogonal projection*** operators). **Table 3** shows the standard matrices for the orthogonal projections onto the coordinate axes in R^2, and **Table 4** shows the standard matrices for the orthogonal projections onto the coordinate planes in R^3.

TABLE 3

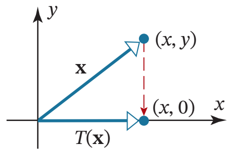
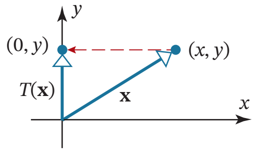
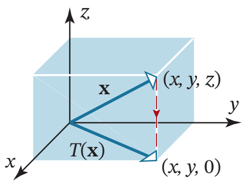
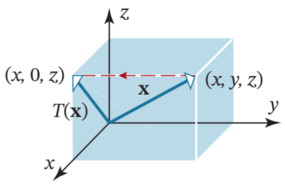
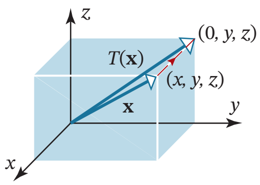
Operator	Illustration	Images of $\mathbf{e}_1$ and $\mathbf{e}_2$	Standard Matrix
Orthogonal projection onto the x-axis $T(x, y) = (x, 0)$		$T(\mathbf{e}_1) = T(1, 0) = (1, 0)$ $T(\mathbf{e}_2) = T(0, 1) = (0, 0)$	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
Orthogonal projection onto the y-axis $T(x, y) = (0, y)$		$T(\mathbf{e}_1) = T(1, 0) = (0, 0)$ $T(\mathbf{e}_2) = T(0, 1) = (0, 1)$	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

TABLE 4

Operator	Illustration	Images of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$	Standard Matrix
Orthogonal projection onto the xy-plane $T(x, y, z) = (x, y, 0)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (1, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, 1, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, 0)$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
Orthogonal projection onto the xz-plane $T(x, y, z) = (x, 0, z)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (1, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, 0, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, 1)$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Orthogonal projection onto the yz-plane $T(x, y, z) = (0, y, z)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (0, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, 1, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, 1)$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Matrix multiplication is really not needed to accomplish the reflections and projections in these tables, as the results are evident geometrically. For example, although the computation

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

shows that the orthogonal projection of (x, y, z) onto the xz-plane is $(x, 0, z)$, that result is evident from the illustration in Table 4. However, in the next section and subsequently we will study more complicated matrix transformations in which the end results are not evident and matrix multiplication is essential.

Rotation Operators

Matrix operators on R^2 that move points along arcs of circles centered at the origin are called ***rotation operators***. Let us consider how to find the standard matrix for the rotation operator $T: R^2 \to R^2$ that moves points *counterclockwise* about the origin through a positive angle θ. **Figure 1.8.4** shows a typical vector $\mathbf{x}$ in R^2 and its image $T(\mathbf{x})$ under such a rotation. As illustrated in **Figure 1.8.5**, the images of the standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ under a rotation through an angle θ are

$$T(\mathbf{e}_1) = T(1,0) = (\cos\theta, \sin\theta) \quad \text{and} \quad T(\mathbf{e}_2) = T(0,1) = (-\sin\theta, \cos\theta)$$

so it follows from Formula (15) that the standard matrix for T is

$$A = [T(\mathbf{e}_1) \mid T(\mathbf{e}_2)] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

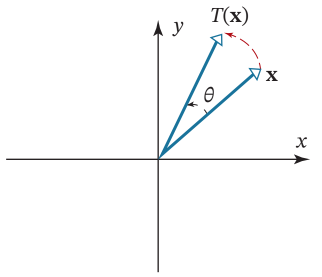


FIGURE 1.8.4

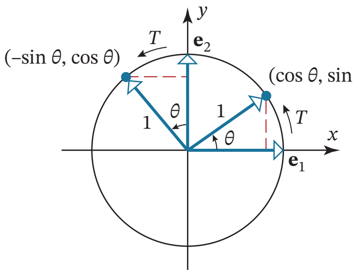


FIGURE 1.8.5

In keeping with common usage we will denote this matrix as

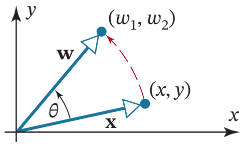
$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{19}$$

and call it the ***rotation matrix*** for R^2. These ideas are summarized in **Table 5**.

In the plane, counterclockwise angles are positive and clockwise angles are negative. The rotation matrix for a *clockwise* rotation of $-\theta$ radians can be obtained by replacing θ by $-\theta$ in (19). After simplification this yields

$$R_{-\theta} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

TABLE 5

Operator	Illustration	Images of $\mathbf{e}_1$ and $\mathbf{e}_2$	Standard Matrix
Counterclockwise rotation about the origin through an angle θ		$T(\mathbf{e}_1) = T(1,0) = (\cos\theta, \sin\theta)$ $T(\mathbf{e}_2) = T(0,1) = (-\sin\theta, \cos\theta)$	$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

EXAMPLE 8 | A Rotation Matrix

Find the image of $\mathbf{x} = (1,1)$ under a rotation of $\pi/6$ radians ($= 30°$) about the origin.

Solution It follows from (19) with $\theta = \pi/6$ that

$$R_{\pi/6}\mathbf{x} = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}-1}{2} \\ \frac{1+\sqrt{3}}{2} \end{bmatrix} \approx \begin{bmatrix} 0.37 \\ 1.37 \end{bmatrix}$$

or in comma-delimited notation, $R_{\pi/6}(1,1) \approx (0.37, 1.37)$.

Concluding Remark

Rotations in R^3 are substantially more complicated than those in R^2 and will be considered later in this text.

Exercise Set 1.8

In Exercises 1–2, find the domain and codomain of the transformation $T_A(\mathbf{x}) = A\mathbf{x}$.

1. **a.** A has size 3×2. **b.** A has size 2×3.
c. A has size 3×3. **d.** A has size 1×6.

2. **a.** A has size 4×5. **b.** A has size 5×4.
c. A has size 4×4. **d.** A has size 3×1.

In Exercises 3–4, find the domain and codomain of the transformation defined by the equations.

3. **a.** $\begin{aligned} w_1 &= 4x_1 + 5x_2 \\ w_2 &= x_1 - 8x_2 \end{aligned}$ **b.** $\begin{aligned} w_1 &= 5x_1 - 7x_2 \\ w_2 &= 6x_1 + x_2 \\ w_3 &= 2x_1 + 3x_2 \end{aligned}$

4. **a.** $\begin{aligned} w_1 &= x_1 - 4x_2 + 8x_3 \\ w_2 &= -x_1 + 4x_2 + 2x_3 \\ w_3 &= -3x_1 + 2x_2 - 5x_3 \end{aligned}$ **b.** $\begin{aligned} w_1 &= 2x_1 + 7x_2 - 4x_3 \\ w_2 &= 4x_1 - 3x_2 + 2x_3 \end{aligned}$

In Exercises 5–6, find the domain and codomain of the transformation defined by the matrix product.

5. **a.** $\begin{bmatrix} 3 & 1 & 2 \\ 6 & 7 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ **b.** $\begin{bmatrix} 2 & -1 \\ 4 & 3 \\ 2 & -5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

6. **a.** $\begin{bmatrix} 6 & 3 \\ -1 & 7 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ **b.** $\begin{bmatrix} 2 & 1 & -6 \\ 3 & 7 & -4 \\ 1 & 0 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

In Exercises 7–8, find the domain and codomain of the transformation T defined by the formula.

7. **a.** $T(x_1, x_2) = (2x_1 - x_2, x_1 + x_2)$
b. $T(x_1, x_2, x_3) = (4x_1 + x_2, x_1 + x_2)$

8. **a.** $T(x_1, x_2, x_3, x_4) = (x_1, x_2)$
b. $T(x_1, x_2, x_3) = (x_1, x_2 - x_3, x_2)$

In Exercises 9–10, find the domain and codomain of the transformation T defined by the formula.

9. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 4x_1 \\ x_1 - x_2 \\ 3x_2 \end{bmatrix}$ **10.** $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_2 \\ x_1 - x_3 \\ 0 \end{bmatrix}$

In Exercises 11–12, find the standard matrix for the transformation defined by the equations.

11. **a.** $\begin{aligned} w_1 &= 2x_1 - 3x_2 + x_3 \\ w_2 &= 3x_1 + 5x_2 - x_3 \end{aligned}$ **b.** $\begin{aligned} w_1 &= 7x_1 + 2x_2 - 8x_3 \\ w_2 &= \quad - x_2 + 5x_3 \\ w_3 &= 4x_1 + 7x_2 - x_3 \end{aligned}$

12. **a.** $\begin{aligned} w_1 &= -x_1 + x_2 \\ w_2 &= 3x_1 - 2x_2 \\ w_3 &= 5x_1 - 7x_2 \end{aligned}$ **b.** $\begin{aligned} w_1 &= x_1 \\ w_2 &= x_1 + x_2 \\ w_3 &= x_1 + x_2 + x_3 \\ w_4 &= x_1 + x_2 + x_3 + x_4 \end{aligned}$

13. Find the standard matrix for the transformation T defined by the formula.

a. $T(x_1, x_2) = (x_2, -x_1, x_1 + 3x_2, x_1 - x_2)$
b. $T(x_1, x_2, x_3, x_4) = (7x_1 + 2x_2 - x_3 + x_4, x_2 + x_3, -x_1)$
c. $T(x_1, x_2, x_3) = (0, 0, 0, 0, 0)$
d. $T(x_1, x_2, x_3, x_4) = (x_4, x_1, x_3, x_2, x_1 - x_3)$

14. Find the standard matrix for the operator T defined by the formula.

a. $T(x_1, x_2) = (2x_1 - x_2, x_1 + x_2)$
b. $T(x_1, x_2) = (x_1, x_2)$
c. $T(x_1, x_2, x_3) = (x_1 + 2x_2 + x_3, x_1 + 5x_2, x_3)$
d. $T(x_1, x_2, x_3) = (4x_1, 7x_2, -8x_3)$

15. Find the standard matrix for the operator $T: R^3 \to R^3$ defined by

$$\begin{aligned} w_1 &= 3x_1 + 5x_2 - x_3 \\ w_2 &= 4x_1 - x_2 + x_3 \\ w_3 &= 3x_1 + 2x_2 - x_3 \end{aligned}$$

and then compute $T(-1, 2, 4)$ by directly substituting in the equations and then by matrix multiplication.

16. Find the standard matrix for the transformation $T: R^4 \to R^2$ defined by

$$\begin{aligned} w_1 &= 2x_1 + 3x_2 - 5x_3 - x_4 \\ w_2 &= x_1 - 5x_2 + 2x_3 - 3x_4 \end{aligned}$$

and then compute $T(1, -1, 2, 4)$ by directly substituting in the equations and then by matrix multiplication.

In Exercises 17–18, find the standard matrix for the transformation and use it to compute $T(\mathbf{x})$. Check your result by substituting directly in the formula for T.

17. **a.** $T(x_1, x_2) = (-x_1 + x_2, x_2)$; $\mathbf{x} = (-1, 4)$
b. $T(x_1, x_2, x_3) = (2x_1 - x_2 + x_3, x_2 + x_3, 0)$; $\mathbf{x} = (2, 1, -3)$

18. **a.** $T(x_1, x_2) = (2x_1 - x_2, x_1 + x_2)$; $\mathbf{x} = (-2, 2)$
b. $T(x_1, x_2, x_3) = (x_1, x_2 - x_3, x_2)$; $\mathbf{x} = (1, 0, 5)$

In Exercises 19–20, find $T_A(\mathbf{x})$, and express your answer in matrix form.

19. **a.** $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$

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

20. **a.** $A = \begin{bmatrix} -2 & 1 & 4 \\ 3 & 5 & 7 \\ 6 & 0 & -1 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

b. $A = \begin{bmatrix} -1 & 1 \\ 2 & 4 \\ 7 & 8 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

In Exercises 21–22, use Theorem 1.8.2 to show that T is a matrix transformation.

21. a. $T(x, y) = (2x + y, x - y)$

b. $T(x_1, x_2, x_3) = (x_1, x_3, x_1 + x_2)$

22. a. $T(x, y, z) = (x + y, y + z, x)$

b. $T(x_1, x_2) = (x_2, x_1)$

In Exercises 23–24, use Theorem 1.8.2 to show that T is not a matrix transformation.

23. a. $T(x, y) = (x^2, y)$

b. $T(x, y, z) = (x, y, xz)$

24. a. $T(x, y) = (x, y + 1)$

b. $T(x_1, x_2, x_3) = (x_1, x_2, \sqrt{x_3})$

25. A function of the form $f(x) = mx + b$ is commonly called a "linear function" because the graph of $y = mx + b$ is a line. Is f a matrix transformation on R?

26. Show that $T(x, y) = (0, 0)$ defines a matrix operator on R^2 but $T(x, y) = (1, 1)$ does not.

In Exercises 27–28, the images of the standard basis vectors for R^3 are given for a linear transformation $T: R^3 \to R^3$. Find the standard matrix for the transformation, and find $T(\mathbf{x})$.

27. $T(\mathbf{e}_1) = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, \; T(\mathbf{e}_2) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \; T(\mathbf{e}_3) = \begin{bmatrix} 4 \\ -3 \\ -1 \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$

28. $T(\mathbf{e}_1) = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \; T(\mathbf{e}_2) = \begin{bmatrix} -3 \\ -1 \\ 0 \end{bmatrix}, \; T(\mathbf{e}_3) = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$

29. Use matrix multiplication to find the reflection of $(-1, 2)$ about the

a. x-axis. **b.** y-axis. **c.** line $y = x$.

30. Use matrix multiplication to find the reflection of (a, b) about the

a. x-axis. **b.** y-axis. **c.** line $y = x$.

31. Use matrix multiplication to find the reflection of $(2, -5, 3)$ about the

a. xy-plane. **b.** xz-plane. **c.** yz-plane.

32. Use matrix multiplication to find the reflection of (a, b, c) about the

a. xy-plane. **b.** xz-plane. **c.** yz-plane.

33. Use matrix multiplication to find the orthogonal projection of $(2, -5)$ onto the

a. x-axis. **b.** y-axis.

34. Use matrix multiplication to find the orthogonal projection of (a, b) onto the

a. x-axis. **b.** y-axis.

35. Use matrix multiplication to find the orthogonal projection of $(-2, 1, 3)$ onto the

a. xy-plane. **b.** xz-plane. **c.** yz-plane.

36. Use matrix multiplication to find the orthogonal projection of (a, b, c) onto the

a. xy-plane. **b.** xz-plane. **c.** yz-plane.

37. Use matrix multiplication to find the image of the vector $(3, -4)$ when it is rotated about the origin through an angle of

a. $\theta = 30^\circ$. **b.** $\theta = -60^\circ$.

c. $\theta = 45^\circ$. **d.** $\theta = 90^\circ$.

38. Use matrix multiplication to find the image of the nonzero vector $\mathbf{v} = (v_1, v_2)$ when it is rotated about the origin through

a. a positive angle α. **b.** a negative angle $-\alpha$.

39. Let $T: R^2 \to R^2$ be a linear operator for which the images of the standard basis vectors for R^2 are $T(\mathbf{e}_1) = (a, b)$ and $T(\mathbf{e}_2) = (c, d)$. Find $T(1, 1)$.

40. Let $T_A: R^2 \to R^2$ be multiplication by

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

and let $\mathbf{e}_1$ and $\mathbf{e}_2$ be the standard basis vectors for R^2. Find the following vectors by inspection.

a. $T_A(k\mathbf{e}_1)$ **b.** $T_A(k\mathbf{e}_1 + l\mathbf{e}_2)$

41. Let $T_A: R^3 \to R^3$ be multiplication by

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

and let $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ be the standard basis vectors for R^3. Find the following vectors by inspection.

a. $T_A(\mathbf{e}_1)$, $T_A(\mathbf{e}_2)$, and $T_A(\mathbf{e}_3)$

b. $T_A(\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3)$ **c.** $T_A(7\mathbf{e}_3)$

42. For each orthogonal projection operator in Table 4 use the standard matrix to compute $T(1, 2, 3)$, and convince yourself that your result makes sense geometrically.

43. For each reflection operator in Table 2 use the standard matrix to compute $T(1, 2, 3)$, and convince yourself that your result makes sense geometrically.

44. If multiplication by A rotates a vector $\mathbf{x}$ in the xy-plane through an angle θ, what is the effect of multiplying $\mathbf{x}$ by A^T? Explain your reasoning.

45. Find the standard matrix A for the linear transformation $T: R^2 \to R^2$ for which

$$T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \; T\left(\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} -2 \\ 5 \end{bmatrix}$$

46. Find the standard matrix A for the linear transformation $T: R^3 \to R^3$ for which

$$T\left(\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ -3 \\ 10 \end{bmatrix}, \; T\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix}, \; T\left(\begin{bmatrix} -3 \\ -1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} -5 \\ -11 \\ 7 \end{bmatrix}$$

47. Let $\mathbf{x}_0$ be a nonzero column vector in R^2, and suppose that $T: R^2 \to R^2$ is the transformation defined by the formula $T(\mathbf{x}) = \mathbf{x}_0 + R_\theta \mathbf{x}$, where R_θ is the standard matrix of the rotation of R^2 about the origin through the angle θ. Give a geometric description of this transformation. Is it a matrix transformation? Explain.

48. In each part of the accompanying figure, find the standard matrix for the pictured operator.

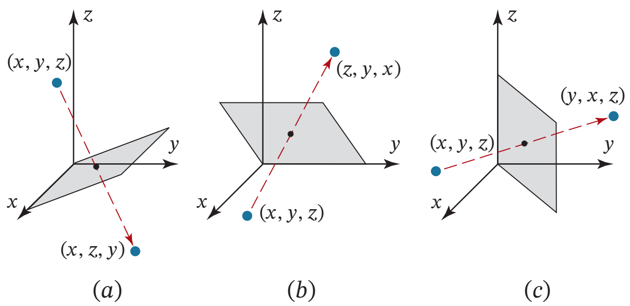


FIGURE Ex-48

49. In a sentence, describe the geometric effect of multiplying a vector $\mathbf{x}$ by the matrix

$$A = \begin{bmatrix} \cos^2\theta - \sin^2\theta & -2\sin\theta\cos\theta \\ 2\sin\theta\cos\theta & \cos^2\theta - \sin^2\theta \end{bmatrix}$$

Working with Proofs

50. a. Prove: If $T: R^n \to R^m$ is a matrix transformation, then $T(\mathbf{0}) = \mathbf{0}$; that is, T maps the zero vector in R^n into the zero vector in R^m.

b. The converse of this is not true. Find an example of a mapping $T: R^n \to R^m$ for which $T(\mathbf{0}) = \mathbf{0}$ but which is not a matrix transformation.

True-False Exercises

TF. In parts **(a)–(g)** determine whether the statement is true or false, and justify your answer.

a. If A is a 2×3 matrix, then the domain of the transformation T_A is R^2.

b. If A is an $m \times n$ matrix, then the codomain of the transformation T_A is R^n.

c. There is at least one linear transformation $T: R^n \to R^m$ for which $T(2\mathbf{x}) = 4T(\mathbf{x})$ for some vector $\mathbf{x}$ in R^n.

d. There are linear transformations from R^n to R^m that are not matrix transformations.

e. If $T_A: R^n \to R^n$ and if $T_A(\mathbf{x}) = \mathbf{0}$ for every vector $\mathbf{x}$ in R^n, then A is the $n \times n$ zero matrix.

f. There is only one matrix transformation $T: R^n \to R^m$ such that $T(-\mathbf{x}) = -T(\mathbf{x})$ for every vector $\mathbf{x}$ in R^n.

g. If $\mathbf{b}$ is a nonzero vector in R^n, then $T(\mathbf{x}) = \mathbf{x} + \mathbf{b}$ is a matrix operator on R^n.

1.9 Compositions of Matrix Transformations

In this section we will discuss the analogs of matrix multiplication and matrix inversion for matrix transformations, and we illustrate those ideas with familiar geometric operations such as rotations, reflections, and projections in the plane. One of the by-products of our work on compositions will be an explanation of why matrix multiplication was defined in such an unusual way.

Compositions of Matrix Transformations

Simply stated, the "composition" of matrix transformations is the process of first applying a matrix transformation to a vector and then applying another matrix transformation to the image vector. For example, suppose that T_A is a matrix transformation from R^n to R^k and T_B is a matrix transformation from R^k to R^m. If $\mathbf{x}$ is a vector in R^n, then T_A maps this vector into a vector $T_A(\mathbf{x})$ in R^k, and T_B, in turn, maps that vector into the vector $T_B(T_A(\mathbf{x}))$ in R^m. This process creates a transformation directly from R^n to R^m that we call the ***composition of T_B with T_A*** and which we denote by the symbol

$$T_B \circ T_A$$

which is read "T_B circle T_A." As illustrated in **Figure 1.9.1**, the transformation T_A in the formula is performed first; that is,

$$\boxed{(T_B \circ T_A)(\mathbf{x}) = T_B(T_A(\mathbf{x}))} \tag{1}$$

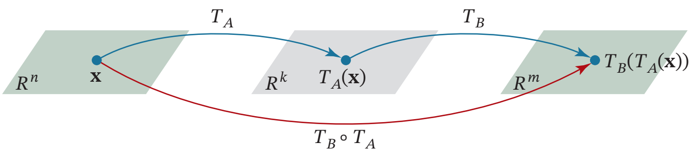


FIGURE 1.9.1

In the introduction to this section we promised to explain why matrix multiplication was defined in such an unusual way. The following theorem does that by showing that our definition of matrix multiplication is precisely what is required to ensure that the composition of two matrix transformations has the same effect as the transformation that results when the underlying matrices are multiplied.

Theorem 1.9.1

If $T_A: R^n \to R^k$ and $T_B: R^k \to R^m$ are matrix transformations, then $T_B \circ T_A$ is also a matrix transformation and

$$T_B \circ T_A = T_{BA} \tag{2}$$

Proof First we will show that $T_B \circ T_A$ is a linear transformation, thereby establishing that it is a matrix transformation by Theorem 1.8.3. Then we will show that the standard matrix for this transformation is BA to complete the proof.

To prove that $T_B \circ T_A$ is linear we must show that it has the additivity and homogeneity properties stated in Theorem 1.8.2. For this purpose, let $\mathbf{x}$ and $\mathbf{y}$ be vectors in R^n and observe that

$$\begin{aligned}
(T_B \circ T_A)(\mathbf{x}+\mathbf{y}) &= T_B(T_A(\mathbf{x}+\mathbf{y})) \\
&= T_B(T_A(\mathbf{x}) + T_A(\mathbf{y})) && [\textbf{because } \boldsymbol{T_A} \textbf{ is linear}] \\
&= T_B(T_A(\mathbf{x})) + T_B(T_A(\mathbf{y})) && [\textbf{because } \boldsymbol{T_B} \textbf{ is linear}] \\
&= (T_B \circ T_A)(\mathbf{x}) + (T_B \circ T_A)(\mathbf{y})
\end{aligned}$$

which proves additivity. Moreover,

$$\begin{aligned}
(T_B \circ T_A)(k\mathbf{x}) &= T_B(T_A(k\mathbf{x})) \\
&= T_B(kT_A(\mathbf{x})) && [\textbf{because } \boldsymbol{T_A} \textbf{ is linear}] \\
&= kT_B(T_A(\mathbf{x})) && [\textbf{because } \boldsymbol{T_B} \textbf{ is linear}] \\
&= k(T_B \circ T_A)(\mathbf{x})
\end{aligned}$$

which proves homogeneity and establishes that $T_B \circ T_A$ is a matrix transformation. Thus, there is an $m \times n$ matrix C such that

$$T_B \circ T_A = T_C \tag{3}$$

To find the appropriate matrix C that satisfies equation (3), observe that

$$T_C(\mathbf{x}) = (T_B \circ T_A)(\mathbf{x}) = T_B(T_A(\mathbf{x})) = T_B(A\mathbf{x}) = B(A\mathbf{x}) = (BA)\mathbf{x} = T_{BA}(\mathbf{x})$$

It now follows from Theorem 1.8.4 that $C = BA$. ■

EXAMPLE 1 | The Standard Matrix for a Composition

Let $T_1: R^3 \to R^2$ and $T_2: R^2 \to R^3$ be the linear transformations given by

$$T_1(x, y, z) = (x + 2y, x + 2z - y)$$

and

$$T_2(x, y) = (3x + y, x, x - 2y)$$

Find the standard matrices for $T_2 \circ T_1$ and $T_1 \circ T_2$.

Solution The standard basis vectors for R^3 are $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$, and $\mathbf{e}_3 = (0, 0, 1)$. From which it follows that

$$T_1(\mathbf{e}_1) = (1,\ 1), \quad T_1(\mathbf{e}_2) = (2, -1) \quad \text{and} \quad T_1(\mathbf{e}_3) = (0, 2)$$

Thus

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

is the standard matrix for T_1. Similarly, the standard basis vectors for R^2 are $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$, so

$$T_2(\mathbf{e}_1) = (3, 1, 1) \quad \text{and} \quad T_2(\mathbf{e}_2) = (1, 0, 2)$$

Thus

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

is the standard matrix for T_2. Applying equation (3), the standard matrix for $T_2 \circ T_1$ is

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

and the standard matrix for $T_1 \circ T_2$ is

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

Commutativity of Matrix Transformations

Since it is *not* generally true that $AB = BA$, it is also *not* generally true that $T_{AB} = T_{BA}$, so in general

$$T_A \circ T_B \neq T_B \circ T_A$$

Thus, *composition of matrix transformations is not commutative.* In those special cases where equality holds, we say that T_A and T_B ***commute***. Note, for example, that the linear transformations in Example 1 do not commute, since $AB \neq BA$.

EXAMPLE 2 | Composition Is Not Commutative

Let $T_A\colon R^2 \to R^2$ be the reflection about the line $y = x$, and let $T_B\colon R^2 \to R^2$ be the orthogonal projection onto the y-axis. **Figure 1.9.2** illustrates graphically that $T_A \circ T_B$ and $T_B \circ T_A$ have different effects on a vector $\mathbf{x}$. This same conclusion can be reached by showing that the standard matrices for T_A and T_B do not commute:

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

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

so $AB \neq BA$.

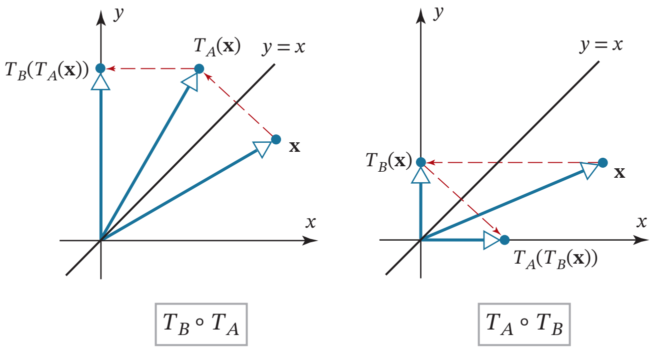


FIGURE 1.9.2

EXAMPLE 3 | Composition of Rotations Is Commutative

It is evident geometrically that the effect of rotating a vector about the origin through an angle θ_1 and then rotating the resulting vector through an angle θ_2 has the same effect as first rotating through the angle θ_2 and then rotating through the angle θ_1 since in both cases the original vector has been rotated through a total angle of $\theta = \theta_1 + \theta_2 = \theta_2 + \theta_1$. This suggests that the matrix transformations $T_{A_1}: R^2 \to R^2$ and $T_{A_2}: R^2 \to R^2$ that rotate vectors about the origin through the angles θ_1 and θ_2, respectively, should commute; that is

$$T_{A_1} \circ T_{A_2} = T_{A_2} \circ T_{A_1}$$

or equivalently

$$T_{A_1A_2} = T_{A_2A_1}$$

To verify that this is so, we need only show that $A_1A_2 = A_2A_1$. But from Table 5 of Section 1.8 we know that

$$A_1 = \begin{bmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{bmatrix}$$

so (with the help of some basic trigonometric identities) it follows that

$$\begin{aligned}
A_1A_2 &= \begin{bmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{bmatrix}\begin{bmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{bmatrix} \\
&= \begin{bmatrix} \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 & -(\cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2) \\ \sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2 & -\sin\theta_1\sin\theta_2 + \cos\theta_1\cos\theta_2 \end{bmatrix} \\
&= \begin{bmatrix} \cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2) \\ \sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2) \end{bmatrix} = \begin{bmatrix} \cos(\theta_2+\theta_1) & -\sin(\theta_2+\theta_1) \\ \sin(\theta_2+\theta_1) & \cos(\theta_2+\theta_1) \end{bmatrix} \\
&= A_2A_1
\end{aligned}$$

Using the notation R_θ for a rotation of R^2 about the origin through an angle θ, the computation in Example 3 shows that

$$R_{\theta_1}R_{\theta_2} = R_{\theta_1+\theta_2}$$

EXAMPLE 4 | Composition of Two Reflections

Let $T_1: R^2 \to R^2$ be the reflection about the y-axis, and let $T_2: R^2 \to R^2$ be the reflection about the x-axis. In this case $T_1 \circ T_2$ and $T_2 \circ T_1$ are the same; both map every vector $\mathbf{x} = (x, y)$ into its negative $-\mathbf{x} = (-x, -y)$ (as evidenced by the following computation and Figure 1.9.3):

$$\begin{aligned}
(T_1 \circ T_2)(x, y) &= T_1(x, -y) = (-x, -y) \\
(T_2 \circ T_1)(x, y) &= T_2(-x, y) = (-x, -y)
\end{aligned}$$

The equality of $T_1 \circ T_2$ and $T_2 \circ T_1$ can also be deduced by showing that the standard matrices for T_1 and T_2 commute. For this purpose let the standard matrices for these transformations be A_1 and A_2, respectively. Then it follows from Table 1 of Section 1.8 that

$$\begin{aligned}
A_1A_2 &= \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \\
A_2A_1 &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}
\end{aligned}$$

We see from **Figure 1.9.3** that the composition $T_1T_2(\mathbf{x}) = T_2T_1(\mathbf{x})$ has the net effect of rotating the vector $\mathbf{x}$ through an angle of $\pi/2$ $(=180°)$, thereby reflecting that vector through the origin into the vector $-\mathbf{x}$. We call the linear operator $T(\mathbf{x}) = -\mathbf{x}$ the ***reflection about the origin.***

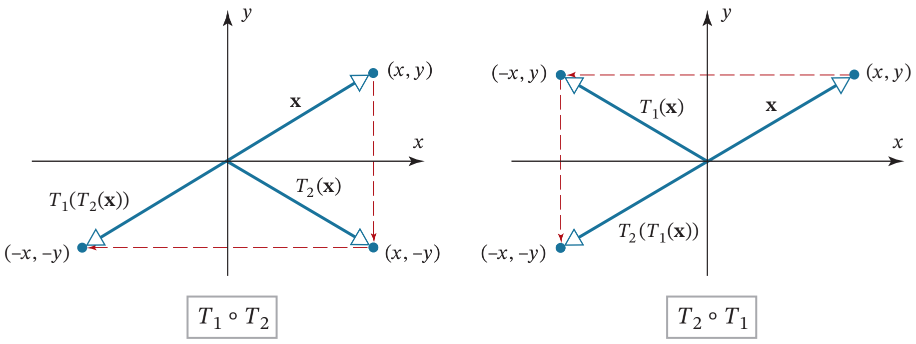


FIGURE 1.9.3

Compositions can be defined for any finite succession of matrix transformations whose domains and ranges have the appropriate dimensions. For example, to extend Formula (3) to three factors, consider the matrix transformations

$$T_A: R^n \to R^k, \quad T_B: R^k \to R^l, T_C: R^l \to R^m$$

We define the composition $(T_C \circ T_B \circ T_A): R^n \to R^m$ by

$$(T_C \circ T_B \circ T_A)(\mathbf{x}) = T_C(T_B(T_A(\mathbf{x})))$$

As above, it can be shown that this is a matrix transformation whose standard matrix is CBA and that

$$\boxed{T_C \circ T_B \circ T_A = T_{CBA}} \tag{4}$$

EXAMPLE 5 | Composition of Three Matrix Transformations

Find the image of a vector

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

under the matrix transformation that first rotates $\mathbf{x}$ about the origin through an angle of $\pi/6$, then reflects the resulting vector about the line $y = x$, and then projects that vector orthogonally onto the y-axis.

Solution Let A, B, and C be the standard matrices for the rotation, the reflection, and the orthogonal projection, respectively. Then from Tables 1, 3, and 5 of Section 1.8 these matrices are

$$A = \begin{bmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

The three transformations in the stated succession can be viewed as the composition

$$T_C \circ T_B \circ T_A = T_{CBA}$$

whose standard matrix is

$$\begin{aligned} CBA &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ \cos(\pi/6) & -\sin(\pi/6) \end{bmatrix} \end{aligned}$$

Thus, the image of the vector $\mathbf{x}$ expressed as a column vector is

$$\begin{bmatrix} 0 & 0 \\ \cos(\pi/6) & -\sin(\pi/6) \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \sqrt{3}/2 & -1/2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ (\sqrt{3}/2)x - (1/2)y \end{bmatrix}$$

Invertibility of Matrix Operators

If $T_A: R^n \to R^n$ is a matrix operator whose standard matrix A is invertible, then we say that T_A is ***invertible***, and we define the ***inverse*** of T_A as

$$\boxed{T_A^{-1} = T_{A^{-1}}} \tag{5}$$

or restated in words, *the inverse of multiplication by A is multiplication by the inverse of A.* Thus, by definition, the standard matrix for T_A^{-1} is A^{-1}, from which it follows that

$$T_A^{-1} \circ T_A = T_{A^{-1}} \circ T_A = T_{A^{-1}A} = T_I$$

It follows from this that for any vector $\mathbf{x}$ in R^n

$$\left(T_A^{-1} \circ T_A\right)(\mathbf{x}) = T_I(\mathbf{x}) = I\mathbf{x} = \mathbf{x}$$

and similarly that $\left(T_A \circ T_A^{-1}\right)(\mathbf{x}) = \mathbf{x}$. Thus, when T_A and T_A^{-1} are composed in either order they cancel out the effect of one another (**Figure 1.9.4**).

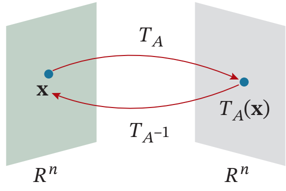


FIGURE 1.9.4

EXAMPLE 6 | Inverse of a Rotation Operator

Let $T: R^2 \to R^2$ be the operator that rotates each vector in R^2 through the angle θ, so the standard matrix for T is

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

It is evident geometrically that to undo the effect of T, one must rotate each vector in R^2 through the angle $-\theta$. But this is precisely what T^{-1} does, since it follows from (5) and Theorem 1.4.5 that the standard matrix for this transformation is

$$R_\theta^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} = R_{-\theta}$$

EXAMPLE 7 | Inverse Transformations from Linear Equations

Consider the operator $T: R^2 \to R^2$ defined by the equations

$$\begin{aligned} w_1 &= 2x_1 + x_2 \\ w_2 &= 3x_1 + 4x_2 \end{aligned}$$

Find $T^{-1}(w_1, w_2)$.

Solution The matrix form of these equations is

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

so the standard matrix for T is

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

This matrix is invertible, and the standard matrix for T^{-1} is

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

Thus

$$A^{-1}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \frac{4}{5} & -\frac{1}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \frac{4}{5}w_1 - \frac{1}{5}w_2 \\ -\frac{3}{5}w_1 + \frac{2}{5}w_2 \end{bmatrix}$$

from which we conclude that

$$T^{-1}(w_1, w_2) = \left(\tfrac{4}{5}w_1 - \tfrac{1}{5}w_2, -\tfrac{3}{5}w_1 + \tfrac{2}{5}w_2\right)$$

Since not every matrix has an inverse, it should not be surprising that the same is true for matrix transformations. As a simple example, consider a transformation $T: R^2 \to R^2$ that projects a vector $\mathbf{x}$ orthogonally onto either the x-axis or the y-axis. You can see in Table 3 of Section 1.8 that the standard matrices for these transformations are not invertible, so in neither case does an invertible matrix A exist to satisfy Equation (5).

Exercise Set 1.9

In Exercises 1–4, determine whether the operators T_1 and T_2 commute; that is, whether $T_1 \circ T_2 = T_2 \circ T_1$.

1. **a.** $T_1: R^2 \to R^2$ is the reflection about the line $y = x$, and $T_2: R^2 \to R^2$ is the orthogonal projection onto the x-axis.

b. $T_1: R^2 \to R^2$ is the reflection about the x-axis, and $T_2: R^2 \to R^2$ is the reflection about the line $y = x$.

2. **a.** $T_1: R^2 \to R^2$ is the orthogonal projection onto the x-axis, and $T_2: R^2 \to R^2$ is the orthogonal projection onto the y-axis.

b. $T_1: R^2 \to R^2$ is the rotation about the origin through an angle of $\pi/4$, and $T_2: R^2 \to R^2$ is the reflection about the y-axis.

3. $T_1: R^3 \to R^3$ is the reflection about the xy-plane and $T_2: R^3 \to R^3$ is the orthogonal projection onto the yz-plane.

4. $T_1: R^3 \to R^3$ is the reflection about the xy-plane and $T_2: R^3 \to R^3$ is given by the formula $T(x, y, z) = (2x, 3y, z)$.

In Exercises 5–6, let T_A and T_B be the operators whose standard matrices are given. Find the standard matrices for $T_B \circ T_A$ and $T_A \circ T_B$.

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

6. $A = \begin{bmatrix} 6 & 3 & -1 \\ 2 & 0 & 1 \\ 4 & -3 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 0 & 4 \\ -1 & 5 & 2 \\ 2 & -3 & 8 \end{bmatrix}$

7. Find the standard matrix for the stated composition in R^2.

a. A rotation of 90°, followed by a reflection about the line $y = x$.

b. An orthogonal projection onto the y-axis, followed by a 45° degree rotation about the origin.

c. A reflection about the x-axis, followed by a rotation about the origin of 60°.

8. Find the standard matrix for the stated composition in R^2.

a. A rotation about the origin of 60°, followed by an orthogonal projection onto the x-axis, followed by a reflection about the line $y = x$.

b. An orthogonal projection onto the x-axis, followed by a rotation about the origin of 45°, followed by a reflection about the y-axis.

c. A rotation about the origin of 15°, followed by a rotation about the origin of 105°, followed by a rotation about the origin of 60°.

9. Find the standard matrix for the stated composition in R^3.

a. A reflection about the yz-plane, followed by an orthogonal projection onto the xz-plane.

b. A reflection about the xy-plane, followed by an orthogonal projection onto the xy-plane.

c. An orthogonal projection onto the xy-plane, followed by a reflection about the yz-plane.

10. Find the standard matrix for the stated composition in R^3.

a. A reflection about the xy-plane, followed by an orthogonal projection onto the xz-plane, followed by the transformation that sends each vector $\mathbf{x}$ to the vector $-\mathbf{x}$.

b. A reflection about the xy-plane, followed by a reflection about the xz-plane, followed by an orthogonal projection onto the yz-plane.

c. An orthogonal projection onto the yz-plane, followed by the transformation that maps each vector $\mathbf{x}$ to the vector $2\mathbf{x}$, followed by a reflection about the xy-plane.

11. Let $T_1(x_1, x_2) = (x_1 + x_2, x_1 - x_2)$ and $T_2(x_1, x_2) = (3x_1, 2x_1 + 4x_2)$.

a. Find the standard matrices for T_1 and T_2.

b. Find the standard matrices for $T_2 \circ T_1$ and $T_1 \circ T_2$.

c. Use the matrices obtained in part (b) to find formulas for $T_1(T_2(x_1, x_2))$ and $T_2(T_1(x_1, x_2))$.

12. Let $T_1(x_1, x_2, x_3) = (4x_1, -2x_1 + x_2, -x_1 - 3x_2)$ and $T_2(x_1, x_2, x_3) = (x_1 + 2x_2, -x_3, 4x_1 - x_3)$.

a. Find the standard matrices for T_1 and T_2.

b. Find the standard matrices for $T_2 \circ T_1$ and $T_1 \circ T_2$.

c. Use the matrices obtained in part (b) to find formulas for $T_1(T_2(x_1, x_2, x_3))$ and $T_2(T_1(x_1, x_2, x_3))$.

13. Let $T_1(x_1, x_2) = (x_1 - x_2, 2x_2 - x_1, 3x_1)$ and $T_2(x_1, x_2, x_3) = (4x_2, x_1 + 2x_2)$.

a. Find the standard matrices for T_1 and T_2.

b. Find the standard matrices for $T_2 \circ T_1$ and $T_1 \circ T_2$.

c. Use the matrices obtained in part (b) to find formulas for $T_1(T_2(x_1, x_2, x_3))$ and $T_2(T_1(x_1, x_2))$.

14. Let $T_1(x_1, x_2, x_3, x_4) = (x_1 + 2x_2 + 3x_3, x_2 - x_4)$ and $T_2(x_1, x_2) = (-x_1, 0, x_1 + x_2, 3x_2)$.

a. Find the standard matrices for T_1 and T_2.

b. Find the standard matrices for $T_2 \circ T_1$ and $T_1 \circ T_2$.

c. Use the matrices obtained in part (b) to find formulas for $T_1(T_2(x_1, x_2))$ and $T_2(T_1(x_1, x_2, x_3, x_4))$.

15. Let $T_1 : R^2 \to R^4$ and $T_2 : R^4 \to R^3$ be given by:
$T_1(x, y) = (y, x, x + y, x - y)$
$T_2(x, y, z, w) = (x + w, y + w, z + w)$.

a. Find the standard matrices for T_1 and T_2.

b. Find the standard matrices for $T_2 \circ T_1$.

c. Explain why $T_1 \circ T_2$ is not defined.

d. Use the matrix found in part (b) to find a formula for $(T_2 \circ T_1)(x, y)$.

16. Let $T_1 : R^2 \to R^3$ and $T_2 : R^3 \to R^4$ be given by:
$T_1(x, y) = (x + 2y, 0, 2x + y)$
$T_2(x, y, z) = (3z, x - y, 3z, y - x)$.

a. Find the standard matrices for T_1 and T_2.

b. Find the standard matrices for $T_2 \circ T_1$.

c. Explain why $T_1 \circ T_2$ is not defined.

d. Use the matrix found in part (b) to find a formula for $(T_2 \circ T_1)(x, y)$.

In Exercises 17–18, express the equations in matrix form, and then use Theorem 1.5.3(c) to determine whether the operator defined by the equations is invertible.

17. a. $\begin{aligned} w_1 &= 8x_1 + 4x_2 \\ w_2 &= 2x_1 + x_2 \end{aligned}$

b. $\begin{aligned} w_1 &= -x_1 + 3x_2 + 2x_3 \\ w_2 &= 2x_1 \phantom{{}+3x_2} + 4x_3 \\ w_3 &= x_1 + 3x_2 + 6x_3 \end{aligned}$

18. a. $\begin{aligned} w_1 &= 2x_1 - 3x_2 \\ w_2 &= 5x_1 + x_2 \end{aligned}$

b. $\begin{aligned} w_1 &= x_1 + 2x_2 + 3x_3 \\ w_2 &= 2x_1 + 5x_2 + 3x_3 \\ w_3 &= x_1 \phantom{{}+5x_2} + 8x_3 \end{aligned}$

19. Determine whether the matrix operator $T : R^2 \to R^2$ defined by the equations is invertible; if so, find the standard matrix for the inverse operator, and find $T^{-1}(w_1, w_2)$.

a. $\begin{aligned} w_1 &= x_1 + 2x_2 \\ w_2 &= -x_1 + x_2 \end{aligned}$

b. $\begin{aligned} w_1 &= 4x_1 - 6x_2 \\ w_2 &= -2x_1 + 3x_2 \end{aligned}$

20. Determine whether the matrix operator $T : R^3 \to R^3$ defined by the equations is invertible; if so, find the standard matrix for the inverse operator, and find $T^{-1}(w_1, w_2, w_3)$.

a. $\begin{aligned} w_1 &= x_1 - 2x_2 + 2x_3 \\ w_2 &= 2x_1 + x_2 + x_3 \\ w_3 &= x_1 + x_2 \end{aligned}$

b. $\begin{aligned} w_1 &= x_1 - 3x_2 + 4x_3 \\ w_2 &= -x_1 + x_2 + x_3 \\ w_3 &= - 2x_2 + 5x_3 \end{aligned}$

In Exercises 21–22, determine whether the matrix operator is invertible. If so, describe in words the effect of its inverse.

21. a. Reflection about the x-axis in R^2.

b. A 60° rotation about the origin in R^2.

c. Orthogonal projection onto the x-axis in R^2.

22. a. Reflection about the line $y = x$.

b. Orthogonal projection onto the y-axis.

c. Reflection about the origin.

In Exercises 23–24, determine whether T_A is invertible. If so, compute $T_A^{-1}(\mathbf{x})$.

23. a. $A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

b. $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

24. a. $A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 2 & 3 & 1 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

b. $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

25. Let $T_A : R^2 \to R^2$ be multiplication by

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

a. What is the geometric effect of applying this transformation to a vector $\mathbf{x}$ in R^2?

b. Express the operator T_A as a composition of two linear operators on R^2.

26. Let $T_A : R^2 \to R^2$ be multiplication by

$$A = \begin{bmatrix} \cos^2\theta - \sin^2\theta & -2\sin\theta\cos\theta \\ 2\sin\theta\cos\theta & \cos^2\theta - \sin^2\theta \end{bmatrix}$$

a. What is the geometric effect of applying this transformation to a vector $\mathbf{x}$ in R^2?

b. Express the operator T_A as a composition of two linear operators on R^2.

Working with Proofs

27. Prove that the matrix transformations T_A and T_B commute if and only if the matrices A and B commute.

28. Let T_A and T_B be matrix operators on R^n. Prove that $T_A \circ T_B$ is invertible if and only if both T_A and T_B are invertible.

29. Prove that the matrix operator T_A on R^n is invertible if and only if for every $\mathbf{b}$ in R^n there exists a unique vector $\mathbf{x}$ in R^n such that $T_A(\mathbf{x}) = \mathbf{b}$.

True-False Exercises

TF. In parts (a)–(g) determine whether the statement is true or false, and justify your answer.

a. If T_A and T_B are matrix operators on R^n, then $T_A(T_B(\mathbf{x})) = T_B(T_A(\mathbf{x}))$ for every vector $\mathbf{x}$ in R^n.

b. If T_A and T_B are matrix operators on R^n and $\mathbf{x}$ is a vector in R^n, then $T_B \circ T_A(\mathbf{x}) = BA\mathbf{x}$

c. A composition of two rotation operators about the origin of R^2 is another rotation about the origin.

d. A composition of two reflection operators in R^2 is another reflection operator.

e. The inverse transformation for a reflection in R^2 about the line $y = x$ is the reflection about the line $y = x$.

f. The inverse transformation for a 90° rotation about the origin in R^2 is a 90° rotation about the origin.

g. The inverse transformation for a reflection about the origin in R^2 is a reflection about the origin.

Working with Technology

T1. **a.** Find the standard matrix for the linear operator on R^2 that performs a counterclockwise rotation of 47° about the origin, followed by a reflection about the y-axis, followed by a counterclockwise rotation of 33° about the origin.

b. Find the image of the point $(1, 1)$ under the operator in part (a).

1.10 Applications of Linear Systems

In this section we will discuss some brief applications of linear systems. These are but a small sample of the wide variety of real-world problems to which our study of linear systems is applicable.

Network Analysis

The concept of a *network* appears in a variety of applications. Loosely stated, a ***network*** is a set of ***branches*** through which something "flows." For example, the branches might be electrical wires through which electricity flows, pipes through which water or oil flows, traffic lanes through which vehicular traffic flows, or economic linkages through which money flows, to name a few possibilities.

In most networks, the branches meet at points, called ***nodes*** or ***junctions***, where the flow divides. For example, in an electrical network, nodes occur where three or more wires join, in a traffic network they occur at street intersections, and in a financial network they occur at banking centers where incoming money is distributed to individuals or other institutions.

In the study of networks, there is generally some numerical measure of the rate at which the medium flows through a branch. For example, the flow rate of electricity is often measured in amperes, the flow rate of water or oil in gallons per minute, the flow rate of traffic in vehicles per hour, and the flow rate of European currency in millions of Euros per day. We will restrict our attention to networks in which there is ***flow conservation*** at each node, by which we mean that *the rate of flow into any node is equal to the rate of flow out of that node.* This ensures that the flow medium does not build up at the nodes and block the free movement of the medium through the network.

A common problem in network analysis is to use known flow rates in certain branches to find the flow rates in all of the branches. Here is an example.

EXAMPLE 1 | Network Analysis Using Linear Systems

Figure 1.10.1 shows a network with four nodes in which the flow rate and direction of flow in certain branches are known. Find the flow rates and directions of flow in the remaining branches.

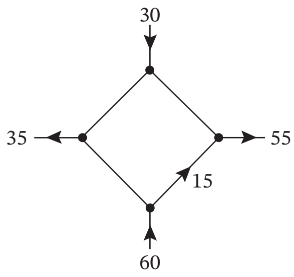


FIGURE 1.10.1

Solution As illustrated in **Figure 1.10.2**, we have assigned arbitrary directions to the unknown flow rates x_1, x_2, and x_3. We need not be concerned if some of the directions are incorrect, since an incorrect direction will be signaled by a negative value for the flow rate when we solve for the unknowns.

It follows from the conservation of flow at node A that

$$x_1 + x_2 = 30$$

Similarly, at the other nodes we have

$$\begin{aligned} x_2 + x_3 &= 35 \quad &(\text{node } B)\\ x_3 + 15 &= 60 \quad &(\text{node } C)\\ x_1 + 15 &= 55 \quad &(\text{node } D) \end{aligned}$$

These four conditions produce the linear system

$$\begin{aligned} x_1 + x_2 \qquad &= 30\\ x_2 + x_3 &= 35\\ x_3 &= 45\\ x_1 \qquad\qquad &= 40 \end{aligned}$$

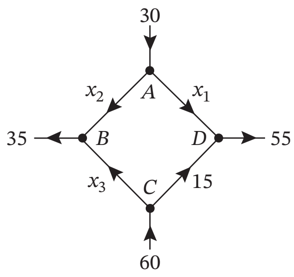


FIGURE 1.10.2

which we can now try to solve for the unknown flow rates. In this particular case the system is sufficiently simple that it can be solved by inspection (work from the bottom up). We leave it for you to confirm that the solution is

$$x_1 = 40, \quad x_2 = -10, \quad x_3 = 45$$

The fact that x_2 is negative tells us that the direction assigned to that flow in Figure 1.10.2 is incorrect; that is, the flow in that branch is *into* node A.

EXAMPLE 2 | Design of Traffic Patterns

The network in **Figure 1.10.3*a*** shows a proposed plan for the traffic flow around a new park that will house the Liberty Bell in Philadelphia, Pennsylvania. The plan calls for a computerized traffic light at the north exit on Fifth Street, and the diagram indicates the average number of vehicles per hour that are expected to flow in and out of the streets that border the complex. All streets are one-way.

(*a*) How many vehicles per hour should the traffic light let through to ensure that the average number of vehicles per hour flowing into the complex is the same as the average number of vehicles flowing out?

(*b*) Assuming that the traffic light has been set to balance the total flow in and out of the complex, what can you say about the average number of vehicles per hour that will flow along the streets that border the complex?

Solution (*a*) If, as indicated in **Figure 1.10.3*b***, we let x denote the number of vehicles per hour that the traffic light must let through, then the total number of vehicles per hour that flow in and out of the complex will be

Flowing in: $500 + 400 + 600 + 200 = 1700$
Flowing out: $x + 700 + 400$

Equating the flows in and out shows that the traffic light should let $x = 600$ vehicles per hour pass through.

Solution (*b*) To avoid traffic congestion, the flow in must equal the flow out at each intersection. For this to happen, the following conditions must be satisfied:

Intersection	Flow In		Flow Out
A	$400 + 600$	$=$	$x_1 + x_2$
B	$x_2 + x_3$	$=$	$400 + x$
C	$500 + 200$	$=$	$x_3 + x_4$
D	$x_1 + x_4$	$=$	700

Thus, with $x = 600$, as computed in part (a), we obtain the following linear system:

$$\begin{aligned} x_1 + x_2 \qquad\qquad &= 1000 \\ x_2 + x_3 \qquad &= 1000 \\ x_3 + x_4 &= 700 \\ x_1 \qquad\qquad + x_4 &= 700 \end{aligned}$$

We leave it for you to show that the system has infinitely many solutions and that these are given by the parametric equations

$$x_1 = 700 - t, \quad x_2 = 300 + t, \quad x_3 = 700 - t, \quad x_4 = t \tag{1}$$

However, the parameter t is not completely arbitrary here, since there are physical constraints to be considered. For example, the average flow rates must be nonnegative since we have assumed the streets to be one-way, and a negative flow rate would indicate a flow in the wrong direction. This being the case, we see from (1) that t can be any real number that satisfies $0 \le t \le 700$, which implies that the average flow rates along the streets will fall in the ranges

$$0 \le x_1 \le 700, \quad 300 \le x_2 \le 1000, \quad 0 \le x_3 \le 700, \quad 0 \le x_4 \le 700$$

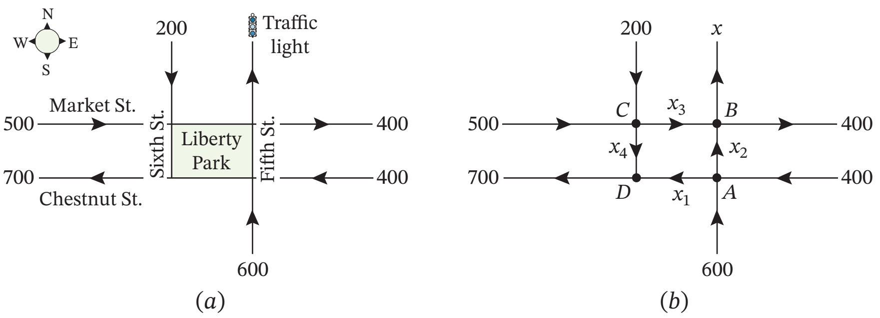


FIGURE 1.10.3

Electrical Circuits

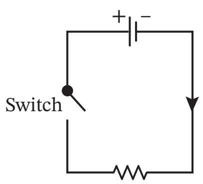


FIGURE 1.10.4

Next we will show how network analysis can be used to analyze electrical circuits consisting of batteries and resistors. A ***battery*** is a source of electric energy, and a ***resistor***, such as a lightbulb, is an element that dissipates electric energy. **Figure 1.10.4** shows a schematic diagram of a circuit with one battery (represented by the symbol ⊣⊢), one resistor (represented by the symbol –\/\/\/–), and a switch. The battery has a ***positive pole*** (+) and a ***negative pole*** (−). When the switch is closed, electrical current is considered to flow from the positive pole of the battery, through the resistor, and back to the negative pole (indicated by the arrowhead in the figure).

Electrical current, which is a flow of electrons through wires, behaves much like the flow of water through pipes. A battery acts like a pump that creates "electrical pressure" to increase the flow rate of electrons, and a resistor acts like a restriction in a pipe that reduces the flow rate of electrons. The technical term for electrical pressure is ***electrical potential***; it is commonly measured in ***volts*** (V). The degree to which a resistor reduces the electrical potential is called its ***resistance*** and is commonly measured in ***ohms*** (Ω). The rate of flow of electrons in a wire is called ***current*** and is commonly measured in ***amperes*** (also called ***amps***) (A). The precise effect of a resistor is given by the following law:

Ohm's Law If a current of I amperes passes through a resistor with a resistance of R ohms, then there is a resulting drop of E volts in electrical potential that is the product of the current and resistance; that is,

$$E = IR$$

A typical electrical network will have multiple batteries and resistors joined by some configuration of wires. A point at which three or more wires in a network are joined is called a ***node*** (or ***junction point***). A ***branch*** is a wire connecting two nodes, and a ***closed loop*** is a succession of connected branches that begin and end at the same node. For example, the electrical network in **Figure 1.10.5** has two nodes and three closed loops—two inner loops and one outer loop. As current flows through an electrical network, it undergoes increases and decreases in electrical potential, called ***voltage rises*** and ***voltage drops***, respectively. The behavior of the current at the nodes and around closed loops is governed by two fundamental laws:

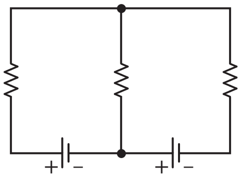


FIGURE 1.10.5

Kirchhoff's Current Law The sum of the currents flowing into any node is equal to the sum of the currents flowing out.

Kirchhoff's Voltage Law In one traversal of any closed loop, the sum of the voltage rises equals the sum of the voltage drops.

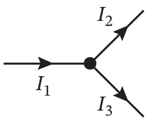
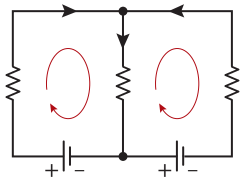


FIGURE 1.10.6

Kirchhoff's current law is a restatement of the principle of flow conservation at a node that was stated for general networks. Thus, for example, the currents at the top node in **Figure 1.10.6** satisfy the equation $I_1 = I_2 + I_3$.

In circuits with multiple loops and batteries there is usually no way to tell in advance which way the currents are flowing, so the usual procedure in circuit analysis is to assign *arbitrary* directions to the current flows in the branches and let the mathematical computations determine whether the assignments are correct. In addition to assigning directions to the current flows, Kirchhoff's voltage law requires a direction of travel for each closed loop. The choice is arbitrary, but for consistency we will always take this direction to be *clockwise* (**Figure 1.10.7**). We also make the following conventions:

- A voltage drop occurs at a resistor if the direction assigned to the current through the resistor is the same as the direction assigned to the loop, and a voltage rise occurs at a resistor if the direction assigned to the current through the resistor is the opposite to that assigned to the loop.
- A voltage rise occurs at a battery if the direction assigned to the loop is from − to + through the battery, and a voltage drop occurs at a battery if the direction assigned to the loop is from + to − through the battery.

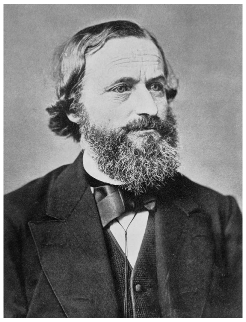


Clockwise closed-loop convention with arbitrary direction assignments to currents in the branches

FIGURE 1.10.7

If you follow these conventions when calculating currents, then those currents whose directions were assigned correctly will have positive values and those whose directions were assigned incorrectly will have negative values.

Historical Note



The German physicist Gustav Kirchhoff was a student of Gauss. His work on Kirchhoff's laws, announced in 1854, was a major advance in the calculation of currents, voltages, and resistances of electrical circuits. Kirchhoff was severely disabled and spent most of his life on crutches or in a wheelchair.

[Image: Courtesy of Library of Congress]

Gustav Kirchhoff
(1824–1887)

EXAMPLE 3 | A Circuit with One Closed Loop

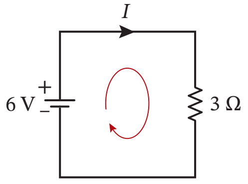


FIGURE 1.10.8

Determine the current I in the circuit shown in Figure 1.10.8.

Solution Since the direction assigned to the current through the resistor is the same as the direction of the loop, there is a voltage drop at the resistor. By Ohm's law this voltage drop is $E = IR = 3I$. Also, since the direction assigned to the loop is from $-$ to $+$ through the battery, there is a voltage rise of 6 volts at the battery. Thus, it follows from Kirchhoff's voltage law that

$$3I = 6$$

from which we conclude that the current is $I = 2$ A. Since I is positive, the direction assigned to the current flow is correct.

EXAMPLE 4 | A Circuit with Three Closed Loops

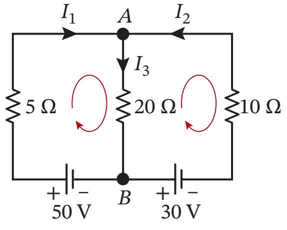


FIGURE 1.10.9

Determine the currents I_1, I_2, and I_3 in the circuit shown in Figure 1.10.9.

Solution Using the assigned directions for the currents, Kirchhoff's current law provides one equation for each node:

Node	Current In		Current Out
A	$I_1 + I_2$	$=$	I_3
B	I_3	$=$	$I_1 + I_2$

However, these equations are really the same, since both can be expressed as

$$I_1 + I_2 - I_3 = 0 \tag{2}$$

To find unique values for the currents we will need two more equations, which we will obtain from Kirchhoff's voltage law. We can see from the network diagram that there are three closed loops, a left inner loop containing the 50 V battery, a right inner loop containing the 30 V battery, and an outer loop that contains both batteries. Thus, Kirchhoff's voltage law will actually produce three equations. With a clockwise traversal of the loops, the voltage rises and drops in these loops are as follows:

	Voltage Rises	Voltage Drops
Left Inside Loop	50	$5I_1 + 20I_3$
Right Inside Loop	$30 + 10I_2 + 20I_3$	0
Outside Loop	$30 + 50 + 10I_2$	$5I_1$

These conditions can be rewritten as

$$\begin{aligned}
5I_1 \qquad\quad + 20I_3 &= 50 \\
10I_2 + 20I_3 &= -30 \\
5I_1 - 10I_2 \qquad\quad &= 80
\end{aligned} \tag{3}$$

However, the last equation is superfluous, since it is the difference of the first two. Thus, if we combine (2) and the first two equations in (3), we obtain the following linear system of three equations in the three unknown currents:

$$\begin{aligned}
I_1 + I_2 - I_3 &= 0 \\
5I_1 \qquad\quad + 20I_3 &= 50 \\
10I_2 + 20I_3 &= -30
\end{aligned}$$

We leave it for you to show that the solution of this system in amps is $I_1 = 6$, $I_2 = -5$, and $I_3 = 1$. The fact that I_2 is negative tells us that the direction of this current is opposite to that indicated in Figure 1.10.9.

Balancing Chemical Equations

Chemical compounds are represented by ***chemical formulas*** that describe the atomic makeup of their molecules. For example, water is composed of two hydrogen atoms and one oxygen atom, so its chemical formula is H_2O; and stable oxygen is composed of two oxygen atoms, so its chemical formula is O_2.

When chemical compounds are combined under the right conditions, the atoms in their molecules rearrange to form new compounds. For example, when methane burns, the methane (CH_4) and stable oxygen (O_2) react to form carbon dioxide (CO_2) and water (H_2O). This is indicated by the ***chemical equation***

$$CH_4 + O_2 \longrightarrow CO_2 + H_2O \tag{4}$$

The molecules to the left of the arrow are called the ***reactants*** and those to the right the ***products***. In this equation the plus signs serve to separate the molecules and are not intended as algebraic operations. However, this equation does not tell the whole story, since it fails to account for the proportions of molecules required for a ***complete reaction*** (no reactants left over). For example, we can see from the right side of (4) that to produce one molecule of carbon dioxide and one molecule of water, one needs *three* oxygen atoms for each carbon atom. However, from the left side of (4) we see that one molecule of methane and one molecule of stable oxygen have only *two* oxygen atoms for each carbon atom. Thus, on the reactant side the ratio of methane to stable oxygen cannot be one-to-one in a complete reaction.

A chemical equation is said to be ***balanced*** if for each type of atom in the reaction, the same number of atoms appears on each side of the arrow. For example, the balanced version of Equation (4) is

$$CH_4 + 2O_2 \longrightarrow CO_2 + 2H_2O \tag{5}$$

by which we mean that one methane molecule combines with two stable oxygen molecules to produce one carbon dioxide molecule and two water molecules. In theory, one could multiply this equation through by any positive integer. For example, multiplying through by 2 yields the balanced chemical equation

$$2CH_4 + 4O_2 \longrightarrow 2CO_2 + 4H_2O$$

However, the standard convention is to use the smallest positive integers that will balance the equation.

Equation (4) is sufficiently simple that it could have been balanced by trial and error, but for more complicated chemical equations we will need a systematic method. There are various methods that can be used, but we will give one that uses systems of linear equations. To illustrate the method let us reexamine Equation (4). To balance this equation we must find positive integers, x_1, x_2, x_3, and x_4 such that

$$x_1\,(CH_4) + x_2\,(O_2) \longrightarrow x_3\,(CO_2) + x_4\,(H_2O) \tag{6}$$

For each of the atoms in the equation, the number of atoms on the left must be equal to the number of atoms on the right. Expressing this in tabular form we have

	Left Side		**Right Side**
Carbon	x_1	$=$	x_3
Hydrogen	$4x_1$	$=$	$2x_4$
Oxygen	$2x_2$	$=$	$2x_3 + x_4$

from which we obtain the homogeneous linear system

$$\begin{aligned} x_1 \qquad - x_3 \qquad &= 0 \\ 4x_1 \qquad\qquad - 2x_4 &= 0 \\ 2x_2 - 2x_3 - x_4 &= 0 \end{aligned}$$

The augmented matrix for this system is

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

We leave it for you to show that the reduced row echelon form of this matrix is

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

from which we conclude that the general solution of the system is

$$x_1 = t/2, \quad x_2 = t, \quad x_3 = t/2, \quad x_4 = t$$

where t is arbitrary. The smallest positive integer values for the unknowns occur when we let $t = 2$, so the equation can be balanced by letting $x_1 = 1, x_2 = 2, x_3 = 1, x_4 = 2$. This agrees with our earlier conclusions, since substituting these values into Equation (6) yields Equation (5).

EXAMPLE 5 | Balancing Chemical Equations Using Linear Systems

Balance the chemical equation

$$\begin{array}{ccccccc} HCl & + & Na_3PO_4 & \longrightarrow & H_3PO_4 & + & NaCl \\ \text{[hydrochloric acid]} & + & \text{[sodium phosphate]} & \longrightarrow & \text{[phosphoric acid]} & + & \text{[sodium chloride]} \end{array}$$

Solution Let x_1, x_2, x_3, and x_4 be positive integers that balance the equation

$$x_1\,(HCl) + x_2\,(Na_3PO_4) \longrightarrow x_3\,(H_3PO_4) + x_4\,(NaCl) \tag{7}$$

Equating the number of atoms of each type on the two sides yields

$$\begin{aligned} 1x_1 &= 3x_3 && \textbf{Hydrogen (H)} \\ 1x_1 &= 1x_4 && \textbf{Chlorine (Cl)} \\ 3x_2 &= 1x_4 && \textbf{Sodium (Na)} \\ 1x_2 &= 1x_3 && \textbf{Phosphorus (P)} \\ 4x_2 &= 4x_3 && \textbf{Oxygen (O)} \end{aligned}$$

from which we obtain the homogeneous linear system

$$\begin{aligned} x_1 \qquad\qquad - 3x_3 \qquad &= 0 \\ x_1 \qquad\qquad\qquad\qquad - x_4 &= 0 \\ 3x_2 \qquad\qquad - x_4 &= 0 \\ x_2 - \ \ x_3 \qquad &= 0 \\ 4x_2 - 4x_3 \qquad &= 0 \end{aligned}$$

We leave it for you to show that the reduced row echelon form of the augmented matrix for this system is

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

from which we conclude that the general solution of the system is

$$x_1 = t, \quad x_2 = t/3, \quad x_3 = t/3, \quad x_4 = t$$

where t is arbitrary. To obtain the smallest positive integers that balance the equation, we let $t = 3$, in which case we obtain $x_1 = 3, x_2 = 1, x_3 = 1$, and $x_4 = 3$. Substituting these values in (7) produces the balanced equation

$$3\text{HCl} + \text{Na}_3\text{PO}_4 \longrightarrow \text{H}_3\text{PO}_4 + 3\text{NaCl}$$

Polynomial Interpolation

An important problem in various applications is to find a polynomial whose graph passes through a specified set of points in the plane; this is called an ***interpolating polynomial*** for the points. The simplest example of such a problem is to find a linear polynomial

$$p(x) = ax + b \tag{8}$$

whose graph passes through two known distinct points, (x_1, y_1) and (x_2, y_2), in the xy-plane (**Figure 1.10.10**). You have probably encountered various methods in analytic geometry for finding the equation of a line through two points, but here we will give a method based on linear systems that can be adapted to general polynomial interpolation.

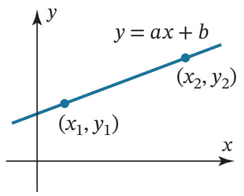


FIGURE 1.10.10

The graph of (8) is the line $y = ax + b$, and for this line to pass through the points (x_1, y_1) and (x_2, y_2), we must have

$$y_1 = ax_1 + b \quad \text{and} \quad y_2 = ax_2 + b$$

Therefore, the unknown coefficients a and b can be obtained by solving the linear system

$$\begin{aligned} ax_1 + b &= y_1 \\ ax_2 + b &= y_2 \end{aligned}$$

We don't need any fancy methods to solve this system—the value of a can be obtained by subtracting the equations to eliminate b, and then the value of a can be substituted into either equation to find b. We leave it as an exercise for you to find a and b and then show that they can be expressed in the form

$$a = \frac{y_2 - y_1}{x_2 - x_1} \quad \text{and} \quad b = \frac{y_1x_2 - y_2x_1}{x_2 - x_1} \tag{9}$$

provided $x_1 \neq x_2$. Thus, for example, the line $y = ax + b$ that passes through the points

$$(2, 1) \quad \text{and} \quad (5, 4)$$

can be obtained by taking $(x_1, y_1) = (2, 1)$ and $(x_2, y_2) = (5, 4)$, in which case (9) yields

$$a = \frac{4-1}{5-2} = 1 \quad \text{and} \quad b = \frac{(1)(5)-(4)(2)}{5-2} = -1$$

Therefore, the equation of the line is

$$y = x - 1$$

(**Figure 1.10.11**).

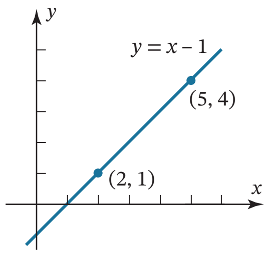


FIGURE 1.10.11

Now let us consider the more general problem of finding a polynomial whose graph passes through n points with distinct x-coordinates

$$(x_1, y_1), \quad (x_2, y_2), \quad (x_3, y_3), \ldots, \quad (x_n, y_n) \tag{10}$$

Since there are n conditions to be satisfied, intuition suggests that we should begin by looking for a polynomial of the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} \tag{11}$$

since a polynomial of this form has n coefficients that are at our disposal to satisfy the n conditions. However, we want to allow for cases where the points may lie on a line or have some other configuration that would make it possible to use a polynomial whose degree is less than $n - 1$; thus, we allow for the possibility that a_{n-1} and other coefficients in (11) may be zero.

The following theorem, which we will not prove, is the basic result on polynomial interpolation.

Theorem 1.10.1

Polynomial Interpolation

Given any n points in the xy-plane that have distinct x-coordinates, there is a unique polynomial of degree $n - 1$ or less whose graph passes through those points.

Let us now consider how we might go about finding the interpolating polynomial (11) whose graph passes through the points in (10). Since the graph of this polynomial is the graph of the equation

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} \tag{12}$$

it follows that the coordinates of the points must satisfy

$$\begin{array}{ccccccccc}
a_0 & + & a_1 x_1 & + & a_2 x_1^2 & + \cdots + & a_{n-1} x_1^{n-1} & = & y_1 \\
a_0 & + & a_1 x_2 & + & a_2 x_2^2 & + \cdots + & a_{n-1} x_2^{n-1} & = & y_2 \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
a_0 & + & a_1 x_n & + & a_2 x_n^2 & + \cdots + & a_{n-1} x_n^{n-1} & = & y_n
\end{array} \tag{13}$$

In these equations the values of x's and y's are assumed to be known, so we can view this as a linear system in the unknowns $a_0, a_1, \ldots, a_{n-1}$. From this point of view the augmented matrix for the system is

$$\begin{bmatrix}
1 & x_1 & x_1^2 & \cdots & x_1^{n-1} & y_1 \\
1 & x_2 & x_2^2 & \cdots & x_2^{n-1} & y_2 \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
1 & x_n & x_n^2 & \cdots & x_n^{n-1} & y_n
\end{bmatrix} \tag{14}$$

and hence the interpolating polynomial can be found by reducing this matrix to reduced row echelon form, say by Gauss-Jordan elimination, as in the following example.

EXAMPLE 6 | Polynomial Interpolation by Gauss–Jordan Elimination

Find a cubic polynomial whose graph passes through the points

$$(1, 3),\quad (2, -2),\quad (3, -5),\quad (4, 0)$$

Solution Since there are four points, we will use an interpolating polynomial of degree $n = 3$. Denote this polynomial by

$$p(x) = a_0 + a_1x + a_2x^2 + a_3x^3$$

and denote the x- and y-coordinates of the given points by

$$x_1 = 1,\quad x_2 = 2,\quad x_3 = 3,\quad x_4 = 4 \quad \text{and} \quad y_1 = 3,\quad y_2 = -2,\quad y_3 = -5,\quad y_4 = 0$$

Thus, it follows from (14) that the augmented matrix for the linear system in the unknowns a_0, a_1, a_2, and a_3 is

$$\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & y_1 \\ 1 & x_2 & x_2^2 & x_2^3 & y_2 \\ 1 & x_3 & x_3^2 & x_3^3 & y_3 \\ 1 & x_4 & x_4^2 & x_4^3 & y_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 3 \\ 1 & 2 & 4 & 8 & -2 \\ 1 & 3 & 9 & 27 & -5 \\ 1 & 4 & 16 & 64 & 0 \end{bmatrix}$$

We leave it for you to confirm that the reduced row echelon form of this matrix is

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

from which it follows that $a_0 = 4, a_1 = 3, a_2 = -5, a_3 = 1$. Thus, the interpolating polynomial is

$$p(x) = 4 + 3x - 5x^2 + x^3$$

The graph of this polynomial and the given points are shown in Figure 1.10.12.

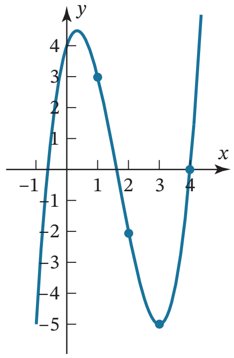


FIGURE 1.10.12

Remark Later we will give a more efficient method for finding interpolating polynomials that is better suited for problems in which the number of data points is large.

EXAMPLE 7 | Approximate Integration

CALCULUS REQUIRED

There is no way to evaluate the integral

$$\int_0^1 \sin\left(\frac{\pi x^2}{2}\right) dx$$

directly since there is no way to express an antiderivative of the integrand in terms of elementary functions. This integral could be approximated by Simpson's rule or some comparable method, but an alternative approach is to approximate the integrand by an interpolating polynomial and integrate the approximating polynomial. For example, let us consider the five points

$$x_0 = 0,\quad x_1 = 0.25,\quad x_2 = 0.5,\quad x_3 = 0.75,\quad x_4 = 1$$

that divide the interval [0, 1] into four equally spaced subintervals (Figure 1.10.13). The values of

$$f(x) = \sin\left(\frac{\pi x^2}{2}\right)$$

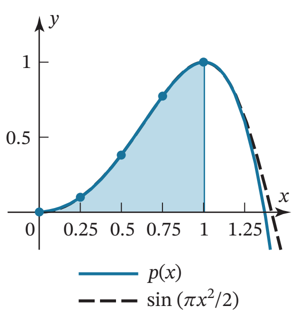


FIGURE 1.10.13

at these points are approximately

$$f(0) = 0, \quad f(0.25) = 0.098017, \quad f(0.5) = 0.382683,$$
$$f(0.75) = 0.77301, \quad f(1) = 1$$

The interpolating polynomial is (verify)

$$p(x) = 0.098796x + 0.762356x^2 + 2.14429x^3 - 2.00544x^4 \tag{15}$$

and

$$\int_0^1 p(x)\,dx \approx 0.438501 \tag{16}$$

As shown in Figure 1.10.13, the graphs of f and p match very closely over the interval $[0, 1]$, so the approximation is quite good.

Exercise Set 1.10

1. The accompanying figure shows a network in which the flow rate and direction of flow in certain branches are known. Find the flow rates and directions of flow in the remaining branches.

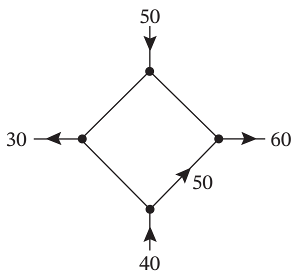


FIGURE Ex-1

2. The accompanying figure shows known flow rates of hydrocarbons into and out of a network of pipes at an oil refinery.

a. Set up a linear system whose solution provides the unknown flow rates.

b. Solve the system for the unknown flow rates.

c. Find the flow rates and directions of flow if $x_4 = 50$ and $x_6 = 0$.

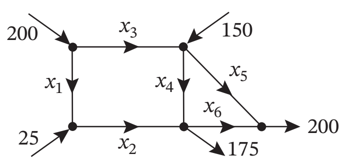


FIGURE Ex-2

3. The accompanying figure shows a network of one-way streets with traffic flowing in the directions indicated. The flow rates along the streets are measured as the average number of vehicles per hour.

a. Set up a linear system whose solution provides the unknown flow rates.

b. Solve the system for the unknown flow rates.

c. If the flow along the road from A to B must be reduced for construction, what is the minimum flow that is required to keep traffic flowing on all roads?

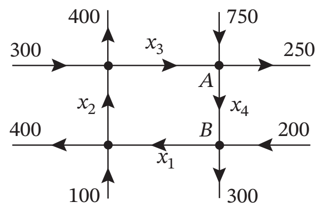


FIGURE Ex-3

4. The accompanying figure shows a network of one-way streets with traffic flowing in the directions indicated. The flow rates along the streets are measured as the average number of vehicles per hour.

a. Set up a linear system whose solution provides the unknown flow rates.

b. Solve the system for the unknown flow rates.

c. Is it possible to close the road from A to B for construction and keep traffic flowing on the other streets? Explain.

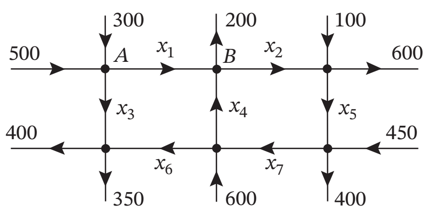
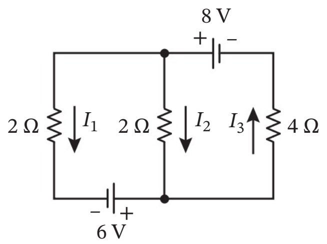


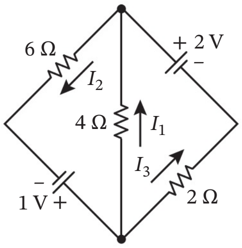
FIGURE Ex-4

In Exercises 5–8, analyze the given electrical circuits by finding the unknown currents.

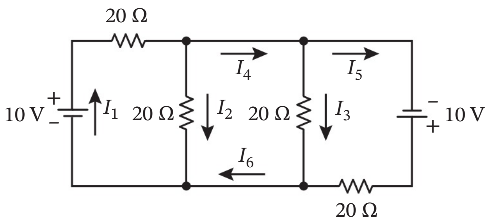
5.



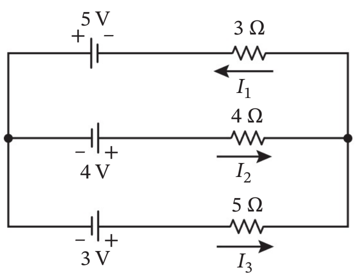
6.



7.



8.



In Exercises 9–12, write a balanced equation for the given chemical reaction.

9. $C_3H_8 + O_2 \rightarrow CO_2 + H_2O$ [propane combustion]

10. $C_6H_{12}O_6 \rightarrow CO_2 + C_2H_5OH$ [fermentation of sugar]

11. $CH_3COF + H_2O \rightarrow CH_3COOH + HF$

12. $CO_2 + H_2O \rightarrow C_6H_{12}O_6 + O_2$ [photosynthesis]

13. Find the quadratic polynomial whose graph passes through the points $(1, 1)$, $(2, 2)$, and $(3, 5)$.

14. Find the quadratic polynomial whose graph passes through the points $(0, 0)$, $(-1, 1)$, and $(1, 1)$.

15. Find the cubic polynomial whose graph passes through the points $(-1, -1)$, $(0, 1)$, $(1, 3)$, $(4, -1)$.

16. The accompanying figure shows the graph of a cubic polynomial. Find the polynomial.

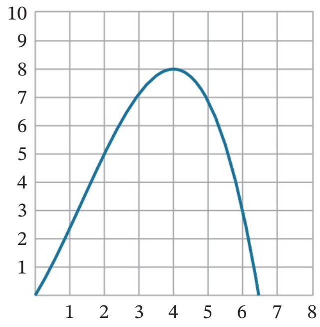


FIGURE Ex-16

17. **a.** Find an equation that represents the family of all second-degree polynomials that pass through the points $(0, 1)$ and $(1, 2)$. [*Hint:* The equation will involve one arbitrary parameter that produces the members of the family when varied.]

b. By hand, or with the help of a graphing utility, sketch four curves in the family.

18. In this section we have selected only a few applications of linear systems. Using the Internet as a search tool, try to find some more real-world applications of such systems. Select one that is of interest to you and write a paragraph about it.

True-False Exercises

TF. In parts (a)–(e) determine whether the statement is true or false, and justify your answer.

a. In any network, the sum of the flows out of a node must equal the sum of the flows into a node.

b. When a current passes through a resistor, there is an increase in the electrical potential in a circuit.

c. Kirchhoff's current law states that the sum of the currents flowing into a node equals the sum of the currents flowing out of the node.

d. A chemical equation is called balanced if the total number of atoms on each side of the equation is the same.

e. Given any n points in the xy-plane, there is a unique polynomial of degree $n - 1$ or less whose graph passes through those points.

Working with Technology

T1. The following table shows the lifting force on an aircraft wing measured in a wind tunnel at various wind velocities. Model the data with an interpolating polynomial of degree 5, and use that polynomial to estimate the lifting force at 2000 ft/s.

Velocity (100 ft/s)	1	2	4	8	16	32
Lifting Force (100 lb)	0	3.12	15.86	33.7	81.5	123.0

T2. (***Calculus required***) Use the method of Example 7 to approximate the integral

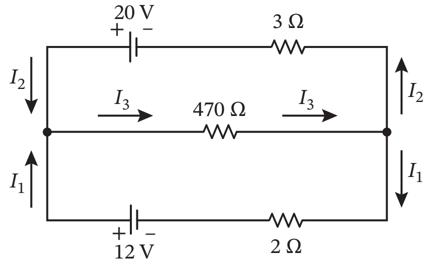
$$\int_0^1 e^{x^2}\,dx$$

by subdividing the interval of integration into five equal parts and using an interpolating polynomial to approximate the integrand. Compare your answer to that obtained using the numerical integration capability of your technology utility.

T3. Use the method of Example 5 to balance the chemical equation

$$Fe_2O_3 + Al \rightarrow Al_2O_3 + Fe$$
$$(Fe = \text{iron}, Al = \text{aluminum}, O = \text{oxygen})$$

T4. Determine the currents in the accompanying circuit.



1.11 Leontief Input-Output Models

In 1973 the economist Wassily Leontief was awarded the Nobel prize for his work on economic modeling in which he used matrix methods to study the relationships among different sectors in an economy. In this section we will discuss some of the ideas developed by Leontief.

Inputs and Outputs in an Economy

One way to analyze an economy is to divide it into ***sectors*** and study how the sectors interact with one another. For example, a simple economy might be divided into three sectors—manufacturing, agriculture, and utilities. Typically, a sector will produce certain ***outputs*** but will require ***inputs*** from the other sectors and itself. For example, the agricultural sector may produce wheat as an output but will require inputs of farm machinery from the manufacturing sector, electrical power from the utilities sector, and food from its own sector to feed its workers. Thus, we can imagine an economy to be a network in which inputs and outputs flow in and out of the sectors; the study of such flows is called ***input-output analysis***. Inputs and outputs are commonly measured in monetary units (dollars or millions of dollars, for example), but other units of measurement are also possible.

The flows between sectors of a real economy are not always obvious. For example, in World War II the United States had a demand for 50,000 new airplanes that required the construction of many new aluminum manufacturing plants. This produced an unexpectedly large demand for certain copper electrical components, which in turn produced a copper shortage. The problem was eventually resolved by using silver borrowed from Fort Knox as a copper substitute. In all likelihood modern input-output analysis would have anticipated the copper shortage.

Most sectors of an economy will produce outputs, but there may exist sectors that consume outputs without producing anything themselves (the consumer market, for example). Those sectors that do not produce outputs are called ***open sectors***. Economies with no open sectors are called ***closed economies***, and economies with one or more open sectors are called ***open economies*** (**Figure 1.11.1**). In this section we will be concerned with economies with one open sector, and our primary goal will be to determine the output levels that are required for the productive sectors to sustain themselves and satisfy the demand of the open sector.

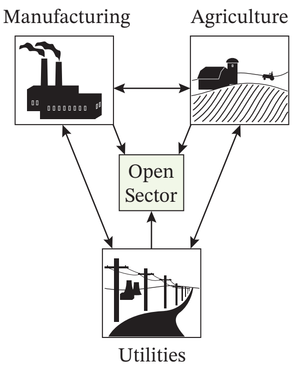


FIGURE 1.11.1

Leontief Model of an Open Economy

Let us consider a simple open economy with one open sector and three product-producing sectors: manufacturing, agriculture, and utilities. Assume that inputs and outputs are measured in dollars and that the inputs required by the productive sectors to produce one dollar's worth of output are in accordance with Table 1.

TABLE 1

		Input Required per Dollar Output		
		Manufacturing	**Agriculture**	**Utilities**
Provider	**Manufacturing**	\$ 0.50	\$ 0.10	\$ 0.10
	Agriculture	\$ 0.20	\$ 0.50	\$ 0.30
	Utilities	\$ 0.10	\$ 0.30	\$ 0.40

Usually, one would suppress the labeling and express this matrix as

$$C = \begin{bmatrix} 0.5 & 0.1 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.3 & 0.4 \end{bmatrix} \tag{1}$$

This is called the ***consumption matrix*** (or sometimes the ***technology matrix***) for the economy. The column vectors

$$\mathbf{c}_1 = \begin{bmatrix} 0.5 \\ 0.2 \\ 0.1 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} 0.1 \\ 0.5 \\ 0.3 \end{bmatrix}, \quad \mathbf{c}_3 = \begin{bmatrix} 0.1 \\ 0.3 \\ 0.4 \end{bmatrix}$$

in C list the inputs required by the manufacturing, agricultural, and utilities sectors, respectively, to produce \$1.00 worth of output. These are called the ***consumption vectors*** of the sectors. For example, $\mathbf{c}_1$ tells us that to produce \$1.00 worth of output the manufacturing sector needs \$0.50 worth of manufacturing output, \$0.20 worth of agricultural output, and \$0.10 worth of utilities output.

What is the economic significance of the row sums of the consumption matrix?

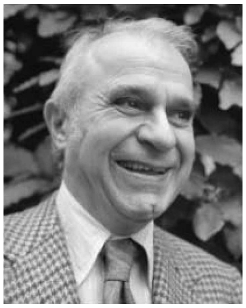
Continuing with the above example, suppose that the open sector wants the economy to supply it manufactured goods, agricultural products, and utilities with dollar values:

d_1 dollars of manufactured goods
d_2 dollars of agricultural products
d_3 dollars of utilities

The column vector $\mathbf{d}$ that has these numbers as successive components is called the ***outside demand vector***. Since the product-producing sectors consume some of their own output, the dollar value of their output must cover their own needs plus the outside demand. Suppose that the dollar values required to do this are

x_1 dollars of manufactured goods
x_2 dollars of agricultural products
x_3 dollars of utilities

Historical Note



Wassily Leontief (1906–1999)

It is somewhat ironic that it was the Russian-born Wassily Leontief who won the Nobel prize in 1973 for pioneering the modern methods for analyzing free-market economies. Leontief was a precocious student who entered the University of Leningrad at age 15. Bothered by the intellectual restrictions of the Soviet system, he was put in jail for anti-Communist activities, after which he headed for the University of Berlin, receiving his Ph.D. there in 1928. He came to the United States in 1931, where he held professorships at Harvard and then New York University.

[Image: © Bettmann/CORBIS]

The column vector $\mathbf{x}$ that has these numbers as successive components is called the ***production vector*** for the economy. For the economy with consumption matrix (1), that portion of the production vector $\mathbf{x}$ that will be consumed by the three productive sectors is

$$\underbrace{x_1\begin{bmatrix}0.5\\0.2\\0.1\end{bmatrix}}_{\substack{\text{Fractions}\\\text{consumed by}\\\text{manufacturing}}} + \underbrace{x_2\begin{bmatrix}0.1\\0.5\\0.3\end{bmatrix}}_{\substack{\text{Fractions}\\\text{consumed by}\\\text{agriculture}}} + \underbrace{x_3\begin{bmatrix}0.1\\0.3\\0.4\end{bmatrix}}_{\substack{\text{Fractions}\\\text{consumed}\\\text{by utilities}}} = \begin{bmatrix}0.5 & 0.1 & 0.1\\0.2 & 0.5 & 0.3\\0.1 & 0.3 & 0.4\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = C\mathbf{x}$$

The vector $C\mathbf{x}$ is called the ***intermediate demand vector*** for the economy. Once the intermediate demand is met, the portion of the production that is left to satisfy the outside demand is $\mathbf{x} - C\mathbf{x}$. Thus, if the outside demand vector is $\mathbf{d}$, then $\mathbf{x}$ must satisfy the equation

$$\underbrace{\mathbf{x}}_{\substack{\text{Amount}\\\text{produced}}} - \underbrace{C\mathbf{x}}_{\substack{\text{Intermediate}\\\text{demand}}} = \underbrace{\mathbf{d}}_{\substack{\text{Outside}\\\text{demand}}}$$

which we will find convenient to rewrite as

$$\boxed{(I - C)\mathbf{x} = \mathbf{d}} \tag{2}$$

The matrix $I - C$ is called the ***Leontief matrix*** and (2) is called the ***Leontief equation***.

EXAMPLE 1 | Satisfying Outside Demand

Consider the economy described in Table 1. Suppose that the open sector has a demand for \$7900 worth of manufacturing products, \$3950 worth of agricultural products, and \$1975 worth of utilities.

(*a*) Can the economy meet this demand?

(*b*) If so, find a production vector $\mathbf{x}$ that will meet it exactly.

Solution The consumption matrix, production vector, and outside demand vector are

$$C = \begin{bmatrix}0.5 & 0.1 & 0.1\\0.2 & 0.5 & 0.3\\0.1 & 0.3 & 0.4\end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix}7900\\3950\\1975\end{bmatrix} \tag{3}$$

To meet the outside demand, the vector $\mathbf{x}$ must satisfy the Leontief equation (2), so the problem reduces to solving the linear system

$$\underbrace{\begin{bmatrix}0.5 & -0.1 & -0.1\\-0.2 & 0.5 & -0.3\\-0.1 & -0.3 & 0.6\end{bmatrix}}_{I-C}\underbrace{\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix}7900\\3950\\1975\end{bmatrix}}_{\mathbf{d}} \tag{4}$$

(if consistent). We leave it for you to show that the reduced row echelon form of the augmented matrix for this system is

$$\left[\begin{array}{ccc|c}1 & 0 & 0 & 27{,}500\\0 & 1 & 0 & 33{,}750\\0 & 0 & 1 & 24{,}750\end{array}\right]$$

This tells us that (4) is consistent, and the economy can satisfy the demand of the open sector exactly by producing \$27,500 worth of manufacturing output, \$33,750 worth of agricultural output, and \$24,750 worth of utilities output.

Productive Open Economies

In the preceding discussion we considered an open economy with three product-producing sectors; the same ideas apply to an open economy with n product-producing sectors. In this case, the consumption matrix, production vector, and outside demand vector have the form

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$$

where all entries are nonnegative and

c_{ij} = the monetary value of the output of the ith sector that is needed by the jth sector to produce one unit of output

x_i = the monetary value of the output of the ith sector

d_i = the monetary value of the output of the ith sector that is required to meet the demand of the open sector

Remark Note that the jth column vector of C contains the monetary values that the jth sector requires of the other sectors to produce one monetary unit of output, and the ith row vector of C contains the monetary values required of the ith sector by the other sectors for each of them to produce one monetary unit of output.

As discussed in our example above, a production vector $\mathbf{x}$ that meets the demand $\mathbf{d}$ of the outside sector must satisfy the Leontief equation

$$(I - C)\mathbf{x} = \mathbf{d}$$

If the matrix $I - C$ is invertible, then this equation has the unique solution

$$\mathbf{x} = (I - C)^{-1}\mathbf{d} \tag{5}$$

for every demand vector $\mathbf{d}$. However, for $\mathbf{x}$ to be a valid production vector it must have nonnegative entries, so the problem of importance in economics is to determine conditions under which the Leontief equation has a solution with nonnegative entries.

It is evident from the form of (5) that if $I - C$ is invertible, and if $(I - C)^{-1}$ has nonnegative entries, then for every demand vector $\mathbf{d}$ the corresponding $\mathbf{x}$ will also have nonnegative entries, and hence will be a valid production vector for the economy. Economies for which $(I - C)^{-1}$ has nonnegative entries are said to be ***productive***. Such economies are desirable because demand can always be met by some level of production. The following theorem, whose proof can be found in many books on economics, gives conditions under which open economies are productive.

Theorem 1.11.1

If C is the consumption matrix for an open economy, and if all of the column sums are less than 1, then the matrix $I - C$ is invertible, the entries of $(I - C)^{-1}$ are nonnegative, and the economy is productive.

Remark The jth column sum of C represents the total dollar value of input that the jth sector requires to produce \$1 of output, so if the jth column sum is less than 1, then the jth sector requires less than \$1 of input to produce \$1 of output; in this case we say that the jth sector is ***profitable***. Thus, Theorem 1.11.1 states that if all product-producing sectors of an open economy are profitable, then the economy is productive. In the exercises we will ask you to show that an open economy is productive if all of the row sums of C are less than 1 (Exercise 11). Thus, an open economy is productive if *either* all of the column sums or all of the row sums of C are less than 1.

EXAMPLE 2 | An Open Economy Whose Sectors Are All Profitable

The column sums of the consumption matrix C in (1) are less than 1, so $(I - C)^{-1}$ exists and has nonnegative entries. Use a calculating utility to confirm this, and use this inverse to solve Equation (4) in Example 1.

Solution We leave it for you to show that

$$(I - C)^{-1} \approx \begin{bmatrix} 2.65823 & 1.13924 & 1.01266 \\ 1.89873 & 3.67089 & 2.15190 \\ 1.39241 & 2.02532 & 2.91139 \end{bmatrix}$$

This matrix has nonnegative entries, and

$$\mathbf{x} = (I - C)^{-1}\mathbf{d} \approx \begin{bmatrix} 2.65823 & 1.13924 & 1.01266 \\ 1.89873 & 3.67089 & 2.15190 \\ 1.39241 & 2.02532 & 2.91139 \end{bmatrix} \begin{bmatrix} 7900 \\ 3950 \\ 1975 \end{bmatrix} \approx \begin{bmatrix} 27{,}500 \\ 33{,}750 \\ 24{,}750 \end{bmatrix}$$

which is consistent with the solution in Example 1.

Exercise Set 1.11

1. An automobile mechanic (M) and a body shop (B) use each other's services. For each \$1.00 of business that M does, it uses \$0.50 of its own services and \$0.25 of B's services, and for each \$1.00 of business that B does it uses \$0.10 of its own services and \$0.25 of M's services.

   a. Construct a consumption matrix for this economy.

   b. How much must M and B each produce to provide customers with \$7000 worth of mechanical work and \$14,000 worth of body work?

2. A simple economy produces food (F) and housing (H). The production of \$1.00 worth of food requires \$0.30 worth of food and \$0.10 worth of housing, and the production of \$1.00 worth of housing requires \$0.20 worth of food and \$0.60 worth of housing.

   a. Construct a consumption matrix for this economy.

   b. What dollar value of food and housing must be produced for the economy to provide consumers \$130,000 worth of food and \$130,000 worth of housing?

3. Consider the open economy described by the accompanying table, where the input is in dollars needed for \$1.00 of output.

   a. Find the consumption matrix for the economy.

   b. Suppose that the open sector has a demand for \$1930 worth of housing, \$3860 worth of food, and \$5790 worth of utilities. Use row reduction to find a production vector that will meet this demand exactly.

TABLE Ex-3

Provider	Input Required per Dollar Output: Housing	Food	Utilities
Housing	\$ 0.10	\$ 0.60	\$ 0.40
Food	\$ 0.30	\$ 0.20	\$ 0.30
Utilities	\$ 0.40	\$ 0.10	\$ 0.20

4. A company produces Web design, software, and networking services. View the company as an open economy described by the accompanying table, where input is in dollars needed for \$1.00 of output.

   a. Find the consumption matrix for the company.

   b. Suppose that the customers (the open sector) have a demand for \$5400 worth of Web design, \$2700 worth of software, and \$900 worth of networking. Use row reduction to find a production vector that will meet this demand exactly.

TABLE Ex-4

Provider	Input Required per Dollar Output: Web Design	Software	Networking
Web Design	\$ 0.40	\$ 0.20	\$ 0.45
Software	\$ 0.30	\$ 0.35	\$ 0.30
Networking	\$ 0.15	\$ 0.10	\$ 0.20

In Exercises 5–6, use matrix inversion to find the production vector $\mathbf{x}$ *that meets the demand* $\mathbf{d}$ *for the consumption matrix* C.

5. $C = \begin{bmatrix} 0.1 & 0.3 \\ 0.5 & 0.4 \end{bmatrix}$; $\mathbf{d} = \begin{bmatrix} 50 \\ 60 \end{bmatrix}$

6. $C = \begin{bmatrix} 0.3 & 0.1 \\ 0.3 & 0.7 \end{bmatrix}$; $\mathbf{d} = \begin{bmatrix} 22 \\ 14 \end{bmatrix}$

7. Consider an open economy with consumption matrix

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

a. Show that the economy can meet a demand of $d_1 = 2$ units from the first sector and $d_2 = 0$ units from the second sector, but it cannot meet a demand of $d_1 = 2$ units from the first sector and $d_2 = 1$ unit from the second sector.

b. Give both a mathematical and an economic explanation of the result in part (a).

8. Consider an open economy with consumption matrix

$$C = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{8} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} \end{bmatrix}$$

If the open sector demands the same dollar value from each product-producing sector, which such sector must produce the greatest dollar value to meet the demand?

9. Consider an open economy with consumption matrix

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & 0 \end{bmatrix}$$

Show that the Leontief equation $\mathbf{x} - C\mathbf{x} = \mathbf{d}$ has a unique solution for every demand vector $\mathbf{d}$ if $c_{21}c_{12} < 1 - c_{11}$.

Working with Proofs

10. a. Consider an open economy with a consumption matrix C whose column sums are less than 1, and let $\mathbf{x}$ be the production vector that satisfies an outside demand $\mathbf{d}$; that is, $(I - C)^{-1}\mathbf{d} = \mathbf{x}$. Let $\mathbf{d}_j$ be the demand vector that is obtained by increasing the jth entry of $\mathbf{d}$ by 1 and leaving the other entries fixed. Prove that the production vector $\mathbf{x}_j$ that meets this demand is

$$\mathbf{x}_j = \mathbf{x} + j\text{th column vector of } (I - C)^{-1}$$

b. In words, what is the economic significance of the jth column vector of $(I - C)^{-1}$? [*Hint:* Look at $\mathbf{x}_j - \mathbf{x}$.]

11. Prove: If C is an $n \times n$ matrix whose entries are nonnegative and whose row sums are less than 1, then $I - C$ is invertible and has nonnegative entries. [*Hint:* $(A^T)^{-1} = (A^{-1})^T$ for any invertible matrix A.]

True-False Exercises

TF. In parts (a)–(e) determine whether the statement is true or false, and justify your answer.

a. Sectors of an economy that produce outputs are called open sectors.

b. A closed economy is an economy that has no open sectors.

c. The rows of a consumption matrix represent the outputs in a sector of an economy.

d. If the column sums of the consumption matrix are all less than 1, then the Leontief matrix is invertible.

e. The Leontief equation relates the production vector for an economy to the outside demand vector.

Working with Technology

T1. The following table describes an open economy with three sectors in which the table entries are the dollar inputs required to produce one dollar of output. The outside demand during a 1-week period if $50,000 of coal, $75,000 of electricity, and $1,250,000 of manufacturing. Determine whether the economy can meet the demand.

Input Required per Dollar Output

Provider	Electricity	Coal	Manufacturing
Electricity	$ 0.1	$ 0.25	$ 0.2
Coal	$ 0.3	$ 0.4	$ 0.5
Manufacturing	$ 0.1	$ 0.15	$ 0.1

Chapter 1 Supplementary Exercises

In Exercises 1–4 the given matrix represents an augmented matrix for a linear system. Write the corresponding set of linear equations for the system, and use Gaussian elimination to solve the linear system. Introduce free parameters as necessary.

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

2. $\begin{bmatrix} 1 & 4 & -1 \\ -2 & -8 & 2 \\ 3 & 12 & -3 \\ 0 & 0 & 0 \end{bmatrix}$

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

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

5. Use Gauss–Jordan elimination to solve for x' and y' in terms of x and y.

$$x = \tfrac{3}{5}x' - \tfrac{4}{5}y'$$
$$y = \tfrac{4}{5}x' + \tfrac{3}{5}y'$$

6. Use Gauss–Jordan elimination to solve for x' and y' in terms of x and y.

$$\begin{aligned} x &= x'\cos\theta - y'\sin\theta \\ y &= x'\sin\theta + y'\cos\theta \end{aligned}$$

7. Find positive integers that satisfy

$$\begin{aligned} x + y + z &= 9 \\ x + 5y + 10z &= 44 \end{aligned}$$

8. A box containing pennies, nickels, and dimes has 13 coins with a total value of 83 cents. How many coins of each type are in the box? Is the economy productive?

9. Let

$$\begin{bmatrix} a & 0 & b & 2 \\ a & a & 4 & 4 \\ 0 & a & 2 & b \end{bmatrix}$$

be the augmented matrix for a linear system. Find for what values of a and b the system has

a. a unique solution.

b. a one-parameter solution.

c. a two-parameter solution. **d.** no solution.

10. For which value(s) of a does the following system have zero solutions? One solution? Infinitely many solutions?

$$\begin{aligned} x_1 + x_2 + x_3 &= 4 \\ x_3 &= 2 \\ (a^2-4)x_3 &= a-2 \end{aligned}$$

11. Find a matrix K such that $AKB = C$ given that

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

$$C = \begin{bmatrix} 8 & 6 & -6 \\ 6 & -1 & 1 \\ -4 & 0 & 0 \end{bmatrix}$$

12. How should the coefficients a, b, and c be chosen so that the system

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

has the solution $x = 1$, $y = -1$, and $z = 2$?

13. In each part, solve the matrix equation for X.

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

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

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

14. Let A be a square matrix.

a. Show that $(I-A)^{-1} = I + A + A^2 + A^3$ if $A^4 = 0$.

b. Show that

$$(I-A)^{-1} = I + A + A^2 + \cdots + A^n$$

if $A^{n+1} = 0$.

15. Find values of a, b, and c such that the graph of the polynomial $p(x) = ax^2 + bx + c$ passes through the points $(1, 2)$, $(-1, 6)$, and $(2, 3)$.

16. ***(Calculus required)*** Find values of a, b, and c such that the graph of $p(x) = ax^2 + bx + c$ passes through the point $(-1, 0)$ and has a horizontal tangent at $(2, -9)$.

17. Let J_n be the $n \times n$ matrix each of whose entries is 1. Show that if $n > 1$, then

$$(I - J_n)^{-1} = I - \frac{1}{n-1}J_n$$

18. Show that if a square matrix A satisfies

$$A^3 + 4A^2 - 2A + 7I = 0$$

then so does A^T.

19. Prove: If B is invertible, then $AB^{-1} = B^{-1}A$ if and only if $AB = BA$.

20. Prove: If A is invertible, then $A + B$ and $I + BA^{-1}$ are both invertible or both not invertible.

21. Prove: If A is an $m \times n$ matrix and B is the $n \times 1$ matrix each of whose entries is $1/n$, then

$$AB = \begin{bmatrix} \bar{r}_1 \\ \bar{r}_2 \\ \vdots \\ \bar{r}_m \end{bmatrix}$$

where $\bar{r}_i$ is the average of the entries in the ith row of A.

22. ***(Calculus required)*** If the entries of the matrix

$$C = \begin{bmatrix} c_{11}(x) & c_{12}(x) & \cdots & c_{1n}(x) \\ c_{21}(x) & c_{22}(x) & \cdots & c_{2n}(x) \\ \vdots & \vdots & & \vdots \\ c_{m1}(x) & c_{m2}(x) & \cdots & c_{mn}(x) \end{bmatrix}$$

are differentiable functions of x, then we define

$$\frac{dC}{dx} = \begin{bmatrix} c'_{11}(x) & c'_{12}(x) & \cdots & c'_{1n}(x) \\ c'_{21}(x) & c'_{22}(x) & \cdots & c'_{2n}(x) \\ \vdots & \vdots & & \vdots \\ c'_{m1}(x) & c'_{m2}(x) & \cdots & c'_{mn}(x) \end{bmatrix}$$

Show that if the entries in A and B are differentiable functions of x and the sizes of the matrices are such that the stated operations can be performed, then

a. $\dfrac{d}{dx}(kA) = k\dfrac{dA}{dx}$

b. $\dfrac{d}{dx}(A+B) = \dfrac{dA}{dx} + \dfrac{dB}{dx}$

c. $\dfrac{d}{dx}(AB) = \dfrac{dA}{dx}B + A\dfrac{dB}{dx}$

23. ***(Calculus required)*** Use part (c) of Exercise 22 to show that

$$\frac{dA^{-1}}{dx} = -A^{-1}\frac{dA}{dx}A^{-1}$$

State all the assumptions you make in obtaining this formula.

24. Assuming that the stated inverses exist, prove the following equalities.

a. $(C^{-1} + D^{-1})^{-1} = C(C + D)^{-1}D$

b. $(I + CD)^{-1}C = C(I + DC)^{-1}$

c. $(C + DD^T)^{-1}D = C^{-1}D(I + D^TC^{-1}D)^{-1}$

Partitioned matrices can be multiplied by the row-column rule just as if the matrix entries were numbers provided that the sizes of all matrices are such that the necessary operations can be performed. Thus, for example, if A is partitioned into a 2×2 matrix and B into a 2×1 matrix, then

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} A_{11}B_1 + A_{12}B_2 \\ A_{21}B_1 + A_{22}B_2 \end{bmatrix} \tag{$*$}$$

provided that the sizes are such that AB, the two sums, and the four products are all defined.

25. Let A and B be the following partitioned matrices.

$$A = \left[\begin{array}{ccc|cc} 1 & 0 & 2 & 1 & 4 \\ 4 & 1 & 0 & 3 & -1 \\ \hline 0 & -3 & 4 & 2 & -2 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

$$B = \left[\begin{array}{cc} 3 & 0 \\ 2 & 1 \\ 4 & -1 \\ \hline 0 & 3 \\ 2 & 5 \end{array}\right] = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$$

a. Confirm that the sizes of all matrices are such that the product AB can be obtained using Formula ($*$).

b. Confirm that the result obtained using Formula ($*$) agrees with that obtained using ordinary matrix multiplication.

26. Suppose that an invertible matrix A is partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

Show that

$$A^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

where

$$B_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}, \qquad B_{12} = -B_{11}A_{12}A_{22}^{-1}$$
$$B_{21} = -A_{22}^{-1}A_{21}B_{11}, \qquad B_{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$$

provided all the inverses in these formulas exist.

27. In the special case where matrix A_{21} in Exercise 26 is zero, the matrix A simplifies to

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

which is said to be in ***block upper triangular form***. Use the result of Exercise 26 to show that in this case

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & A_{22}^{-1} \end{bmatrix}$$

28. A linear system whose coefficient matrix has a pivot position in every row must be consistent. Explain why this must be so.

29. What can you say about the consistency or inconsistency of a linear system of three equations in five unknowns whose coefficient matrix has three pivot columns?

CHAPTER 2

Determinants

CHAPTER CONTENTS

Introduction

In this chapter we will study "determinants" or, more precisely, "determinant functions." Unlike real-valued functions, such as $f(x) = x^2$, that assign a real number to a real variable x, determinant functions assign a real number $f(A)$ to a matrix variable A. Although determinants first arose in the context of solving systems of linear equations, they are rarely used for that purpose in real-world applications. While they can be useful for solving very small linear systems (say, two or three unknowns), our main interest in them stems from the fact that they link together various concepts in linear algebra and provide a useful formula for the inverse of a matrix.

2.1 Determinants by Cofactor Expansion

In this section we will define the notion of a "determinant." This will enable us to develop a specific formula for the inverse of an invertible matrix, whereas up to now we have had only a computational procedure. This, in turn, will eventually provide us with a formula for solutions of certain kinds of linear systems.

Recall from Theorem 1.4.5 that the 2×2 matrix

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

is invertible if and only if $ad - bc \neq 0$ and that the expression $ad - bc$ is called the ***determinant*** of the matrix A. Recall also that this determinant is denoted by writing

Warning It is important to keep in mind that $\det(A)$ is a *number*, whereas A is a *matrix*.

$$\det(A) = ad - bc \quad \text{or} \quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{1}$$

and that the inverse of A can be expressed in terms of the determinant as

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{2}$$

Minors and Cofactors

One of our main goals in this chapter is to obtain an analog of Formula (2) that is applicable to square matrices of *all orders*. For this purpose we will find it convenient to use subscripted entries when writing matrices or determinants. Thus, if we denote a 2×2 matrix as

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

then the two equations in (1) take the form

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \tag{3}$$

In situations where it is inconvenient to assign a name to the matrix, we can express this formula as

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21} \tag{4}$$

There are various methods for defining determinants of higher-order square matrices. In this text, we will use an "inductive definition" by which we mean that the determinant of a square matrix of a given order will be defined in terms of determinants of square matrices of the next lower order. To start the process, let us define the determinant of a 1×1 matrix $[a_{11}]$ as

$$\det[a_{11}] = a_{11} \tag{5}$$

from which it follows that Formula (4) can be expressed as

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \det[a_{11}]\det[a_{22}] - \det[a_{12}]\det[a_{21}]$$

Now that we have established a starting point, we can define determinants of 3×3 matrices in terms of determinants of 2×2 matrices, then determinants of 4×4 matrices in terms of determinants of 3×3 matrices, and so forth, ad infinitum. The following terminology and notation will help to make this inductive process more efficient.

Definition 1

If A is a square matrix, then the ***minor of entry a_{ij}*** is denoted by M_{ij} and is defined to be the determinant of the submatrix that remains after the ith row and jth column are deleted from A. The number $(-1)^{i+j}M_{ij}$ is denoted by C_{ij} and is called the ***cofactor of entry a_{ij}***.

Historical Note

The term *determinant* was first introduced by the German mathematician Carl Friedrich Gauss in 1801 (see p. 16), who used them to "determine" properties of certain kinds of functions. Interestingly, the term *matrix* is derived from a Latin word for "womb" because it was viewed as a container of determinants.

EXAMPLE 1 | Finding Minors and Cofactors

Let

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

Warning We have followed the standard convention of using capital letters to denote minors and cofactors even though they are numbers, not matrices.

The minor of entry a_{11} is

$$M_{11} = \begin{vmatrix} 3 & 1 & -4 \\ 2 & 5 & 6 \\ 1 & 4 & 8 \end{vmatrix} = \begin{vmatrix} 5 & 6 \\ 4 & 8 \end{vmatrix} = 16$$

The cofactor of a_{11} is

$$C_{11} = (-1)^{1+1}M_{11} = M_{11} = 16$$

Similarly, the minor of entry a_{32} is

$$M_{32} = \begin{vmatrix} 3 & 1 & -4 \\ 2 & 5 & 6 \\ 1 & 4 & 8 \end{vmatrix} = \begin{vmatrix} 3 & -4 \\ 2 & 6 \end{vmatrix} = 26$$

The cofactor of a_{32} is

$$C_{32} = (-1)^{3+2}M_{32} = -M_{32} = -26$$

Remark Note that a minor M_{ij} and its corresponding cofactor C_{ij} are either the same or negatives of each other and that the relating sign $(-1)^{i+j}$ is either $+1$ or -1 in accordance with the pattern in the "checkerboard" array

$$\begin{bmatrix} + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

For example,

$$C_{11} = M_{11}, \quad C_{21} = -M_{21}, \quad C_{22} = M_{22}$$

and so forth. Thus, it is never really necessary to calculate $(-1)^{i+j}$ to obtain C_{ij}—you can simply compute the minor M_{ij} and then adjust the sign in accordance with the checkerboard pattern. Try this in Example 1.

EXAMPLE 2 | Cofactor Expansions of a 2×2 Matrix

The checkerboard pattern for a 2×2 matrix $A = [a_{ij}]$ is

$$\begin{bmatrix} + & - \\ - & + \end{bmatrix}$$

so that

$$\begin{aligned} C_{11} &= M_{11} = a_{22} & \qquad C_{12} &= -M_{12} = -a_{21} \\ C_{21} &= -M_{21} = -a_{12} & \qquad C_{22} &= M_{22} = a_{11} \end{aligned}$$

We leave it for you to use Formula (3) to verify that $\det(A)$ can be expressed in terms of cofactors in the following four ways:

$$\begin{aligned} \det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} &= a_{11}C_{11} + a_{12}C_{12} \\ &= a_{21}C_{21} + a_{22}C_{22} \\ &= a_{11}C_{11} + a_{21}C_{21} \\ &= a_{12}C_{12} + a_{22}C_{22} \end{aligned} \tag{6}$$

Each of the last four equations is called a *cofactor expansion* of $\det(A)$. In each cofactor expansion the entries and cofactors all come from the same row or same column of A. For example, in the first equation the entries and cofactors all come from the first row of A, in the second they all come from the second row of A, in the third they all come from the first column of A, and in the fourth they all come from the second column of A.

Historical Note

The term *minor* is apparently due to the English mathematician James Sylvester (see p. 36), who wrote the following in a paper published in 1850: "Now conceive any one line and any one column be struck out, we get . . . a square, one term less in breadth and depth than the original square; and by varying in every possible selection of the line and column excluded, we obtain, supposing the original square to consist of n lines and n columns, n^2 such minor squares, each of which will represent what I term a "First Minor Determinant" relative to the principal or complete determinant."

Definition of a General Determinant

Formula (6) is a special case of the following general result, which we will state without proof.

Theorem 2.1.1

If A is an $n \times n$ matrix, then regardless of which row or column of A is chosen, the number obtained by multiplying the entries in that row or column by the corresponding cofactors and adding the resulting products is always the same.

This result allows us to make the following definition.

Definition 2

If A is an $n \times n$ matrix, then the number obtained by multiplying the entries in any row or column of A by the corresponding cofactors and adding the resulting products is called the ***determinant of A***, and the sums themselves are called ***cofactor expansions of A***. That is,

$$\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} \tag{7}$$

[cofactor expansion along the *j*th column]

and

$$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} \tag{8}$$

[cofactor expansion along the *i*th row]

EXAMPLE 3 | Cofactor Expansion Along the First Row

Find the determinant of the matrix

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

by cofactor expansion along the first row.

Solution

$$\det(A) = \begin{vmatrix} 3 & 1 & 0 \\ -2 & -4 & 3 \\ 5 & 4 & -2 \end{vmatrix} = 3\begin{vmatrix} -4 & 3 \\ 4 & -2 \end{vmatrix} - 1\begin{vmatrix} -2 & 3 \\ 5 & -2 \end{vmatrix} + 0\begin{vmatrix} -2 & -4 \\ 5 & 4 \end{vmatrix}$$

$$= 3(-4) - (1)(-11) + 0 = -1$$

Historical Note

Charles Lutwidge Dodgson (Lewis Carroll) (1832–1898)

Cofactor expansion is not the only method for expressing the determinant of a matrix in terms of determinants of lower order. For example, although it is not well known, the English mathematician Charles Dodgson, who was the author of *Alice's Adventures in Wonderland* and *Through the Looking Glass* under the pen name of Lewis Carroll, invented such a method, called *condensation*. That method has recently been resurrected from obscurity because of its suitability for parallel processing on computers.

[Image: Oscar G. Rejlander/Time & Life Pictures/Getty Images]

EXAMPLE 4 | Cofactor Expansion Along the First Column

Let A be the matrix in Example 3, and evaluate $\det(A)$ by cofactor expansion along the first column of A.

Solution

$$\det(A) = \begin{vmatrix} 3 & 1 & 0 \\ -2 & -4 & 3 \\ 5 & 4 & -2 \end{vmatrix} = 3\begin{vmatrix} -4 & 3 \\ 4 & -2 \end{vmatrix} - (-2)\begin{vmatrix} 1 & 0 \\ 4 & -2 \end{vmatrix} + 5\begin{vmatrix} 1 & 0 \\ -4 & 3 \end{vmatrix}$$

$$= 3(-4) - (-2)(-2) + 5(3) = -1$$

This agrees with the result obtained in Example 3.

Note that in Example 4 we had to compute three cofactors, whereas in Example 3 only two were needed because the third was multiplied by zero. As a rule, the best strategy for cofactor expansion is to expand along a row or column with the most zeros.

EXAMPLE 5 | Smart Choice of Row or Column

If A is the 4×4 matrix

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

then to find $\det(A)$ it will be easiest to use cofactor expansion along the second column, since it has the most zeros:

$$\det(A) = 1 \cdot \begin{vmatrix} 1 & 0 & -1 \\ 1 & -2 & 1 \\ 2 & 0 & 1 \end{vmatrix}$$

For the 3×3 determinant, it will be easiest to use cofactor expansion along its second column, since it has the most zeros:

$$\begin{aligned} \det(A) &= (1)(-2)\begin{vmatrix} 1 & -1 \\ 2 & 1 \end{vmatrix} \\ &= -2(1+2) \\ &= -6 \end{aligned}$$

EXAMPLE 6 | Determinant of a Lower Triangular Matrix

The following computation shows that the determinant of a 4×4 lower triangular matrix is the product of its diagonal entries. Each part of the computation uses a cofactor expansion along the first row.

$$\begin{aligned} \begin{vmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} &= a_{11}\begin{vmatrix} a_{22} & 0 & 0 \\ a_{32} & a_{33} & 0 \\ a_{42} & a_{43} & a_{44} \end{vmatrix} \\ &= a_{11}a_{22}\begin{vmatrix} a_{33} & 0 \\ a_{43} & a_{44} \end{vmatrix} \\ &= a_{11}a_{22}a_{33}|a_{44}| = a_{11}a_{22}a_{33}a_{44} \end{aligned}$$

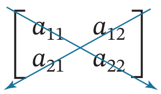
The method illustrated in Example 6 can be easily adapted to prove the following general result.

Theorem 2.1.2

If A is an $n \times n$ triangular matrix (upper triangular, lower triangular, or diagonal), then $\det(A)$ is the product of the entries on the main diagonal of the matrix; that is, $\det(A) = a_{11}a_{22}\cdots a_{nn}$.

A Useful Technique for Evaluating 2×2 and 3×3 Determinants

Determinants of 2×2 and 3×3 matrices can be evaluated very efficiently using the pattern suggested in **Figure 2.1.1**.



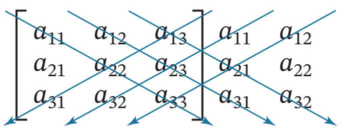


FIGURE 2.1.1

In the 2×2 case, the determinant can be computed by forming the product of the entries on the rightward arrow and subtracting the product of the entries on the leftward arrow. In the 3×3 case we first recopy the first and second columns as shown in the figure, after which we can compute the determinant by summing the products of the entries on the rightward arrows and subtracting the products on the leftward arrows. These procedures execute the computations

Warning The arrow technique works only for determinants of 2×2 and 3×3 matrices. It *does not* work for matrices of size 4×4 or higher.

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

$$\begin{aligned} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} \end{aligned}$$

which agrees with the cofactor expansions along the first row.

EXAMPLE 7 | A Technique for Evaluating 2×2 and 3×3 Determinants

$$\begin{vmatrix} 3 & 1 \\ 4 & -2 \end{vmatrix} = \begin{vmatrix} 3 & 1 \\ 4 & -2 \end{vmatrix} = (3)(-2) - (1)(4) = -10$$

$$\begin{vmatrix} 1 & 2 & 3 \\ -4 & 5 & 6 \\ 7 & -8 & 9 \end{vmatrix} = \left|\begin{array}{ccc|cc} 1 & 2 & 3 & 1 & 2 \\ -4 & 5 & 6 & -4 & 5 \\ 7 & -8 & 9 & 7 & -8 \end{array}\right|$$

$$= [45 + 84 + 96] - [105 - 48 - 72] = 240$$

Exercise Set 2.1

In Exercises **1–2**, *find all the minors and cofactors of the matrix A.*

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

3. Let

$$A = \begin{bmatrix} 4 & -1 & 1 & 6 \\ 0 & 0 & -3 & 3 \\ 4 & 1 & 0 & 14 \\ 4 & 1 & 3 & 2 \end{bmatrix}$$

Find

a. M_{13} and C_{13}. **b.** M_{23} and C_{23}.

c. M_{22} and C_{22}. **d.** M_{21} and C_{21}.

4. Let

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

Find

a. M_{32} and C_{32}. **b.** M_{44} and C_{44}.

c. M_{41} and C_{41}. **d.** M_{24} and C_{24}.

In Exercises **5–8**, *evaluate the determinant of the given matrix. If the matrix is invertible, use Equation (2) to find its inverse.*

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

In Exercises 9–14, use the arrow technique of Figure 2.1.1 to evaluate the determinant.

9. $\begin{vmatrix} a-3 & 5 \\ -3 & a-2 \end{vmatrix}$

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

11. $\begin{vmatrix} -2 & 1 & 4 \\ 3 & 5 & -7 \\ 1 & 6 & 2 \end{vmatrix}$

12. $\begin{vmatrix} -1 & 1 & 2 \\ 3 & 0 & -5 \\ 1 & 7 & 2 \end{vmatrix}$

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

14. $\begin{vmatrix} c & -4 & 3 \\ 2 & 1 & c^2 \\ 4 & c-1 & 2 \end{vmatrix}$

In Exercises 15–18, find all values of λ for which $\det(A) = 0$.

15. $A = \begin{bmatrix} \lambda-2 & 1 \\ -5 & \lambda+4 \end{bmatrix}$

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

17. $A = \begin{bmatrix} \lambda-1 & 0 \\ 2 & \lambda+1 \end{bmatrix}$

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

19. Evaluate the determinant in Exercise 13 by a cofactor expansion along

a. the first row.
b. the first column.
c. the second row.
d. the second column.
e. the third row.
f. the third column.

20. Evaluate the determinant in Exercise 12 by a cofactor expansion along

a. the first row.
b. the first column.
c. the second row.
d. the second column.
e. the third row.
f. the third column.

In Exercises 21–26, evaluate $\det(A)$ by a cofactor expansion along a row or column of your choice.

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

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

23. $A = \begin{bmatrix} 1 & k & k^2 \\ 1 & k & k^2 \\ 1 & k & k^2 \end{bmatrix}$

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

25. $A = \begin{bmatrix} 3 & 3 & 0 & 5 \\ 2 & 2 & 0 & -2 \\ 4 & 1 & -3 & 0 \\ 2 & 10 & 3 & 2 \end{bmatrix}$

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

In Exercises 27–32, evaluate the determinant of the given matrix by inspection.

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

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

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

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

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

32. $\begin{bmatrix} -3 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 40 & 10 & -1 & 0 \\ 100 & 200 & -23 & 3 \end{bmatrix}$

33. In each part, show that the value of the determinant is independent of θ.

a. $\begin{vmatrix} \sin\theta & \cos\theta \\ -\cos\theta & \sin\theta \end{vmatrix}$

b. $\begin{vmatrix} \sin\theta & \cos\theta & 0 \\ -\cos\theta & \sin\theta & 0 \\ \sin\theta-\cos\theta & \sin\theta+\cos\theta & 1 \end{vmatrix}$

34. Show that the matrices

$$A = \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} d & e \\ 0 & f \end{bmatrix}$$

commute if and only if

$$\begin{vmatrix} b & a-c \\ e & d-f \end{vmatrix} = 0$$

35. By inspection, what is the relationship between the following determinants?

$$d_1 = \begin{vmatrix} a & b & c \\ d & 1 & f \\ g & 0 & 1 \end{vmatrix} \quad \text{and} \quad d_2 = \begin{vmatrix} a+\lambda & b & c \\ d & 1 & f \\ g & 0 & 1 \end{vmatrix}$$

36. Show that

$$\det(A) = \frac{1}{2}\begin{vmatrix} \operatorname{tr}(A) & 1 \\ \operatorname{tr}(A^2) & \operatorname{tr}(A) \end{vmatrix}$$

for every 2×2 matrix A.

37. What can you say about an nth-order determinant all of whose entries are 1? Explain.

38. What is the maximum number of zeros that a 3×3 matrix can have without having a zero determinant? Explain.

39. Explain why the determinant of a matrix with integer entries must be an integer.

Working with Proofs

40. Prove that (x_1, y_1), (x_2, y_2), and (x_3, y_3) are collinear points if and only if

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = 0$$

41. Prove that the equation of the line through the distinct points (a_1, b_1) and (a_2, b_2) can be written as

$$\begin{vmatrix} x & y & 1 \\ a_1 & b_1 & 1 \\ a_2 & b_2 & 1 \end{vmatrix} = 0$$

42. Prove that if A is upper triangular and B_{ij} is the matrix that results when the ith row and jth column of A are deleted, then B_{ij} is upper triangular if $i < j$.

43. A matrix in which the entries in each row (or in each column) form a geometric progression starting with 1 is called a ***Vandermonde matrix*** in honor of the French medical doctor, mathematician, and musician Alexandre-Théophile Vandermonde (February 28, 1735–January 1, 1796). Here are two examples.

$$V = \begin{bmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} 1 & a & a^2 & a^3 \\ 1 & b & b^2 & b^3 \\ 1 & c & c^2 & c^3 \\ 1 & d & d^2 & d^3 \end{bmatrix}$$

Vandermonde matrices arise in a variety of applications, such as polynomial interpolation (see Formula (14) and Example 6 of Section 1.10). Use cofactor expansion to prove that

$$\begin{vmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{vmatrix} = (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)$$

True-False Exercises

TF. In parts (a)–(j) determine whether the statement is true or false, and justify your answer.

a. The determinant of the 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $ad + bc$.

b. Two square matrices that have the same determinant must have the same size.

c. The minor M_{ij} is the same as the cofactor C_{ij} if $i + j$ is even.

d. If A is a 3×3 symmetric matrix, then $C_{ij} = C_{ji}$ for all i and j.

e. The number obtained by a cofactor expansion of a matrix A is independent of the row or column chosen for the expansion.

f. If A is a square matrix whose minors are all zero, then $\det(A) = 0$.

g. The determinant of a lower triangular matrix is the sum of the entries along the main diagonal.

h. For every square matrix A and every scalar c, it is true that $\det(cA) = c\det(A)$.

i. For all square matrices A and B, it is true that

$$\det(A + B) = \det(A) + \det(B)$$

j. For every 2×2 matrix A it is true that

$$\det(A^2) = (\det(A))^2$$

Working with Technology

T1. a. Use the determinant capability of your technology utility to find the determinant of the matrix

$$A = \begin{bmatrix} 4.2 & -1.3 & 1.1 & 6.0 \\ 0.0 & 0.0 & -3.2 & 3.4 \\ 4.5 & 1.3 & 0.0 & 14.8 \\ 4.7 & 1.0 & 3.4 & 2.3 \end{bmatrix}$$

b. Compare the result obtained in part (a) to that obtained by a cofactor expansion along the second row of A.

T2. Let A^n be the $n \times n$ matrix with 2's along the main diagonal, 1's along the diagonal lines immediately above and below the main diagonal, and zeros everywhere else. Make a conjecture about the relationship between n and $\det(A_n)$.

2.2 Evaluating Determinants by Row Reduction

In this section we will show how to evaluate a determinant by reducing the associated matrix to row echelon form. In general, this method requires less computation than cofactor expansion and hence is the method of choice for large matrices.

A Basic Theorem

We begin with a fundamental theorem that will lead us to an efficient procedure for evaluating the determinant of a square matrix of any size.

Theorem 2.2.1

Let A be a square matrix. If A has a row of zeros or a column of zeros, then $\det(A) = 0$.

Proof Since the determinant of A can be found by a cofactor expansion along any row or column, we can use the row or column of zeros. Thus, if we let $C_1, C_2, \dots, C_n$ denote the cofactors of A along that row or column, then it follows from Formula (7) or (8) in Section 2.1 that

$$\det(A) = 0 \cdot C_1 + 0 \cdot C_2 + \cdots + 0 \cdot C_n = 0 \quad \blacksquare$$

The following useful theorem relates the determinant of a matrix and the determinant of its transpose.

Theorem 2.2.2

Let A be a square matrix. Then $\det(A) = \det(A^T)$.

Because transposing a matrix changes its columns to rows and its rows to columns, almost every theorem about the rows of a determinant has a companion version about columns, and vice versa.

Proof Since transposing a matrix changes its columns to rows and its rows to columns, the cofactor expansion of A along any row is the same as the cofactor expansion of A^T along the corresponding column. Thus, both have the same determinant. $\blacksquare$

Elementary Row Operations

The next theorem shows how an elementary row operation on a square matrix affects the value of its determinant. In place of a formal proof we have provided a table to illustrate the ideas in the 3×3 case (see **Table 1**).

Theorem 2.2.3

Let A be an $n \times n$ matrix.

(a) If B is the matrix that results when a single row or single column of A is multiplied by a scalar k, then $\det(B) = k\det(A)$.

(b) If B is the matrix that results when two rows or two columns of A are interchanged, then $\det(B) = -\det(A)$.

(c) If B is the matrix that results when a multiple of one row of A is added to another or when a multiple of one column is added to another, then $\det(B) = \det(A)$.

TABLE 1

Relationship	Operation
$\begin{vmatrix} ka_{11} & ka_{12} & ka_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = k \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ $\det(B) = k\det(A)$	In the matrix B the first row of A was multiplied by k.
$\begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = - \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ $\det(B) = -\det(A)$	In the matrix B the first and second rows of A were interchanged.
$\begin{vmatrix} a_{11}+ka_{21} & a_{12}+ka_{22} & a_{13}+ka_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ $\det(B) = \det(A)$	In the matrix B a multiple of the second row of A was added to the first row.

The first panel of Table 1 shows that you can bring a common factor from any row (column) of a determinant through the determinant sign. This is a slightly different way of thinking about part (a) of Theorem 2.2.3.

We will verify the first equation in Table 1 and leave the other two for you. To start, note that the determinants on the two sides of the equation differ only in the first row, so these determinants have the same cofactors, C_{11}, C_{12}, C_{13}, along that row (since those cofactors depend only on the entries in the *second* two rows). Thus, expanding the left side by cofactors along the first row yields

$$\begin{vmatrix} ka_{11} & ka_{12} & ka_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = ka_{11}C_{11} + ka_{12}C_{12} + ka_{13}C_{13}$$

$$= k(a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13})$$

$$= k\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

Elementary Matrices

It will be useful to consider the special case of Theorem 2.2.3 in which $A = I_n$ is the $n \times n$ identity matrix and E (rather than B) denotes the elementary matrix that results when the row operation is performed on I_n. In this special case Theorem 2.2.3 implies the following result.

Theorem 2.2.4

Let E be an $n \times n$ elementary matrix.

(*a*) If E results from multiplying a row of I_n by a nonzero number k, then $\det(E) = k$.

(*b*) If E results from interchanging two rows of I_n, then $\det(E) = -1$.

(*c*) If E results from adding a multiple of one row of I_n to another, then $\det(E) = 1$.

Observe that the determinant of an elementary matrix cannot be zero.

EXAMPLE 1 | Determinants of Elementary Matrices

The following determinants of elementary matrices, which are evaluated by inspection, illustrate Theorem 2.2.4.

$$\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 3, \qquad \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix} = -1, \qquad \begin{vmatrix} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 1$$

The second row of I_4 was multiplied by 3. **The first and last rows of I_4 were interchanged.** **7 times the last row of I_4 was added to the first row.**

Matrices with Proportional Rows or Columns

If a square matrix A has two proportional rows, then a row of zeros can be introduced by adding a suitable multiple of one of those rows to the other. Similarly for columns. But adding a multiple of one row or column to another does not change the determinant, so from Theorem 2.2.1, we must have $\det(A) = 0$. This proves the following theorem.

Theorem 2.2.5

If A is a square matrix with two proportional rows or two proportional columns, then $\det(A) = 0$.

EXAMPLE 2 | Proportional Rows or Columns

Each of the following matrices has two proportional rows or columns; thus, each has a determinant of zero.

$$\begin{bmatrix} -1 & 4 \\ -2 & 8 \end{bmatrix}, \quad \begin{bmatrix} 1 & -2 & 7 \\ -4 & 8 & 5 \\ 2 & -4 & 3 \end{bmatrix}, \quad \begin{bmatrix} 3 & -1 & 4 & -5 \\ 6 & -2 & 5 & 2 \\ 5 & 8 & 1 & 4 \\ -9 & 3 & -12 & 15 \end{bmatrix}$$

Evaluating Determinants by Row Reduction

We will now give a method for evaluating determinants that involves substantially less computation than cofactor expansion. The idea of the method is to reduce the given matrix to upper triangular form by elementary row operations, then compute the determinant of the upper triangular matrix (an easy computation), and then relate that determinant to that of the original matrix. Here is an example.

Even with today's fastest computers it would take millions of years to calculate a 25×25 determinant by cofactor expansion, so methods based on row reduction are often used for large determinants. For determinants of small size (such as those in this text), cofactor expansion is often a reasonable choice.

EXAMPLE 3 | Using Row Reduction to Evaluate a Determinant

Evaluate $\det(A)$ where

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

Solution We will reduce A to row echelon form (which is upper triangular) and then apply Theorem 2.1.2.

$$\det(A) = \begin{vmatrix} 0 & 1 & 5 \\ 3 & -6 & 9 \\ 2 & 6 & 1 \end{vmatrix} = -\begin{vmatrix} 3 & -6 & 9 \\ 0 & 1 & 5 \\ 2 & 6 & 1 \end{vmatrix}$$

The first and second rows of A were interchanged.

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

A common factor of 3 from the first row was taken through the determinant sign.

$$= -3\begin{vmatrix} 1 & -2 & 3 \\ 0 & 1 & 5 \\ 0 & 10 & -5 \end{vmatrix}$$ ← -2 times the first row was added to the third row.

$$= -3\begin{vmatrix} 1 & -2 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & -55 \end{vmatrix}$$ ← -10 times the second row was added to the third row.

$$= (-3)(-55)\begin{vmatrix} 1 & -2 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{vmatrix}$$ ← A common factor of -55 from the last row was taken through the determinant sign.

$$= (-3)(-55)(1) = 165$$

EXAMPLE 4 | Using Column Operations to Evaluate a Determinant

Compute the determinant of

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

Solution This determinant could be computed as above by using elementary row operations to reduce A to row echelon form, but we can put A in lower triangular form in one step by adding -3 times the first column to the fourth to obtain

$$\det(A) = \det\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 7 & 0 & 0 \\ 0 & 6 & 3 & 0 \\ 7 & 3 & 1 & -26 \end{bmatrix} = (1)(7)(3)(-26) = -546$$

Example 4 points out that it is always wise to keep an eye open for column operations that can shorten computations.

Cofactor expansion and row or column operations can sometimes be used in combination to provide an effective method for evaluating determinants. The following example illustrates this idea.

EXAMPLE 5 | Row Operations and Cofactor Expansion

Evaluate $\det(A)$ where

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

Solution By adding suitable multiples of the second row to the remaining rows, we obtain

$$\det(A) = \begin{vmatrix} 0 & -1 & 1 & 3 \\ 1 & 2 & -1 & 1 \\ 0 & 0 & 3 & 3 \\ 0 & 1 & 8 & 0 \end{vmatrix}$$

$$= -\begin{vmatrix} -1 & 1 & 3 \\ 0 & 3 & 3 \\ 1 & 8 & 0 \end{vmatrix} \quad \text{Cofactor expansion along the first column}$$

$$= -\begin{vmatrix} -1 & 1 & 3 \\ 0 & 3 & 3 \\ 0 & 9 & 3 \end{vmatrix} \quad \text{We added the first row to the third row.}$$

$$= -(-1)\begin{vmatrix} 3 & 3 \\ 9 & 3 \end{vmatrix} \quad \text{Cofactor expansion along the first column}$$

$$= -18$$

Exercise Set 2.2

In Exercises 1–4, verify that $\det(A) = \det(A^T)$.

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

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

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

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

In Exercises 5–8, find the determinant of the given elementary matrix by inspection.

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

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

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

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

In Exercises 9–14, evaluate the determinant of the matrix by first reducing the matrix to row echelon form and then using some combination of row operations and cofactor expansion.

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

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

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

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

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

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

In Exercises 15–22, evaluate the determinant, given that

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = -6$$

15. $\begin{vmatrix} d & e & f \\ g & h & i \\ a & b & c \end{vmatrix}$

16. $\begin{vmatrix} g & h & i \\ d & e & f \\ a & b & c \end{vmatrix}$

17. $\begin{vmatrix} 3a & 3b & 3c \\ -d & -e & -f \\ 4g & 4h & 4i \end{vmatrix}$

18. $\begin{vmatrix} a+d & b+e & c+f \\ -d & -e & -f \\ g & h & i \end{vmatrix}$

19. $\begin{vmatrix} a+g & b+h & c+i \\ d & e & f \\ g & h & i \end{vmatrix}$

20. $\begin{vmatrix} a & b & c \\ 2d & 2e & 2f \\ g+3a & h+3b & i+3c \end{vmatrix}$

21. $\begin{vmatrix} -3a & -3b & -3c \\ d & e & f \\ g-4d & h-4e & i-4f \end{vmatrix}$

22. $\begin{vmatrix} a & b & c \\ d & e & f \\ 2a & 2b & 2c \end{vmatrix}$

23. Use row reduction to show that

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix} = (b-a)(c-a)(c-b)$$

24. Verify the formulas in parts (a) and (b) and then make a conjecture about a general result of which these results are special cases.

a. $\det\begin{bmatrix} 0 & 0 & a_{13} \\ 0 & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = -a_{13}a_{22}a_{31}$

b. $\det\begin{bmatrix} 0 & 0 & 0 & a_{14} \\ 0 & 0 & a_{23} & a_{24} \\ 0 & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} = a_{14}a_{23}a_{32}a_{41}$

In Exercises 25–28, confirm the identities without evaluating any of the determinants directly.

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

26. $\begin{vmatrix} a_1+b_1t & a_2+b_2t & a_3+b_3t \\ a_1t+b_1 & a_2t+b_2 & a_3t+b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = (1-t^2)\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$

27. $\begin{vmatrix} a_1+b_1 & a_1-b_1 & c_1 \\ a_2+b_2 & a_2-b_2 & c_2 \\ a_3+b_3 & a_3-b_3 & c_3 \end{vmatrix} = -2\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$

28. $\begin{vmatrix} a_1 & b_1+ta_1 & c_1+rb_1+sa_1 \\ a_2 & b_2+ta_2 & c_2+rb_2+sa_2 \\ a_3 & b_3+ta_3 & c_3+rb_3+sa_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$

In Exercises 29–30, show that $\det(A)=0$ *without directly evaluating the determinant.*

29. $A = \begin{bmatrix} -2 & 8 & 1 & 4 \\ 3 & 2 & 5 & 1 \\ 1 & 10 & 6 & 5 \\ 4 & -6 & 4 & -3 \end{bmatrix}$

30. $A = \begin{bmatrix} -4 & 1 & 1 & 1 & 1 \\ 1 & -4 & 1 & 1 & 1 \\ 1 & 1 & -4 & 1 & 1 \\ 1 & 1 & 1 & -4 & 1 \\ 1 & 1 & 1 & 1 & -4 \end{bmatrix}$

It can be proved that if a square matrix M is partitioned into ***block triangular form*** *as*

$$M = \begin{bmatrix} A & 0 \\ C & B \end{bmatrix} \quad \text{or} \quad M = \begin{bmatrix} A & C \\ 0 & B \end{bmatrix}$$

in which A and B are square, then $\det(M) = \det(A)\det(B)$. *Use this result to compute the determinants of the matrices in Exercises 31 and 32.*

31. $M = \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 8 & 6 & -9 \\ 2 & 5 & 0 & 4 & 7 & 5 \\ -1 & 3 & 2 & 6 & 9 & -2 \\ \hline 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & -3 & 8 & -4 \end{array}\right]$

32. $M = \left[\begin{array}{ccc|cc} 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 2 \\ 2 & 0 & 0 & 0 & 1 \end{array}\right]$

33. Let A be an $n \times n$ matrix, and let B be the matrix that results when the rows of A are written in reverse order. State a theorem that describes how $\det(A)$ and $\det(B)$ are related.

34. Find the determinant of the following matrix.

$$\begin{bmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{bmatrix}$$

True-False Exercises

TF. In parts (a)–(f) determine whether the statement is true or false, and justify your answer.

a. If A is a 4×4 matrix and B is obtained from A by interchanging the first two rows and then interchanging the last two rows, then $\det(B) = \det(A)$.

b. If A is a 3×3 matrix and B is obtained from A by multiplying the first column by 4 and multiplying the third column by $\frac{3}{4}$, then $\det(B) = 3\det(A)$.

c. If A is a 3×3 matrix and B is obtained from A by adding 5 times the first row to each of the second and third rows, then $\det(B) = 25\det(A)$.

d. If A is an $n \times n$ matrix and B is obtained from A by multiplying each row of A by its row number, then

$$\det(B) = \frac{n(n+1)}{2}\det(A)$$

e. If A is a square matrix with two identical columns, then $\det(A) = 0$.

f. If the sum of the second and fourth row vectors of a 6×6 matrix A is equal to the last row vector, then $\det(A) = 0$.

Working with Technology

T1. Find the determinant of

$$A = \begin{bmatrix} 4.2 & -1.3 & 1.1 & 6.0 \\ 0.0 & 0.0 & -3.2 & 3.4 \\ 4.5 & 1.3 & 0.0 & 14.8 \\ 4.7 & 1.0 & 3.4 & 2.3 \end{bmatrix}$$

by reducing the matrix to reduced row echelon form, and compare the result obtained in this way to that obtained in Exercise T1 of Section 2.1.

2.3 Properties of Determinants; Cramer's Rule

In this section we will develop some fundamental properties of matrices, and we will use these results to derive a formula for the inverse of an invertible matrix and formulas for the solutions of certain kinds of linear systems.

Basic Properties of Determinants

Suppose that A and B are $n \times n$ matrices and k is any scalar. We begin by considering possible relationships among $\det(A)$, $\det(B)$, and

$$\det(kA), \quad \det(A+B), \quad \text{and} \quad \det(AB)$$

Since a common factor of any row of a matrix can be moved through the determinant sign, and since each of the n rows in kA has a common factor of k, it follows that

$$\boxed{\det(kA) = k^n \det(A)} \tag{1}$$

For example,

$$\begin{vmatrix} ka_{11} & ka_{12} & ka_{13} \\ ka_{21} & ka_{22} & ka_{23} \\ ka_{31} & ka_{32} & ka_{33} \end{vmatrix} = k^3 \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

Unfortunately, no simple relationship exists among $\det(A)$, $\det(B)$, and $\det(A+B)$. In particular, $\det(A+B)$ will usually *not* be equal to $\det(A)+\det(B)$. The following example illustrates this fact.

EXAMPLE 1 | $\det(A+B) \neq \det(A)+\det(B)$

Consider

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

We have $\det(A) = 1$, $\det(B) = 8$, and $\det(A+B) = 23$; thus

$$\det(A+B) \neq \det(A) + \det(B)$$

In spite of the previous example, there is a useful relationship concerning sums of determinants that is applicable when the matrices involved are the same except for *one* row or column. For example, consider the following two matrices that differ only in the second row:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} a_{11} & a_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

Calculating the determinants of A and B, we obtain

$$\begin{aligned} \det(A) + \det(B) &= (a_{11}a_{22} - a_{12}a_{21}) + (a_{11}b_{22} - a_{12}b_{21}) \\ &= a_{11}(a_{22}+b_{22}) - a_{12}(a_{21}+b_{21}) \\ &= \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{bmatrix} \end{aligned}$$

Thus

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \det \begin{bmatrix} a_{11} & a_{12} \\ b_{21} & b_{22} \end{bmatrix} = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{bmatrix}$$

This is a special case of the following general result.

Theorem 2.3.1

Let A, B, and C be $n \times n$ matrices that differ only in a single row, say the rth, and assume that the rth row of C can be obtained by adding corresponding entries in the rth rows of A and B. Then

$$\det(C) = \det(A) + \det(B)$$

The same result holds for columns.

EXAMPLE 2 | Sums of Determinants

We leave it to you to confirm the following equality by evaluating the determinants.

$$\det\begin{bmatrix} 1 & 7 & 5 \\ 2 & 0 & 3 \\ 1+0 & 4+1 & 7+(-1) \end{bmatrix} = \det\begin{bmatrix} 1 & 7 & 5 \\ 2 & 0 & 3 \\ 1 & 4 & 7 \end{bmatrix} + \det\begin{bmatrix} 1 & 7 & 5 \\ 2 & 0 & 3 \\ 0 & 1 & -1 \end{bmatrix}$$

Determinant of a Matrix Product

Considering the complexity of the formulas for determinants and matrix multiplication, it would seem unlikely that a simple relationship should exist between them. This is what makes the simplicity of our next result so surprising. We will show that if A and B are square matrices of the same size, then

$$\det(AB) = \det(A)\det(B) \tag{2}$$

The proof of this theorem is fairly intricate, so we will have to develop some preliminary results first. We begin with the special case of (2) in which A is an elementary matrix. Because this special case is only a prelude to (2), we call it a lemma.

Lemma 2.3.2

If B is an $n \times n$ matrix and E is an $n \times n$ elementary matrix, then

$$\det(EB) = \det(E)\det(B)$$

Proof We will consider three cases, each in accordance with the row operation that produces the matrix E.

Case 1 If E results from multiplying a row of I_n by k, then by Theorem 1.5.1, EB results from B by multiplying the corresponding row by k; so from Theorem 2.2.3(a) we have

$$\det(EB) = k\det(B)$$

But from Theorem 2.2.4(a) we have $\det(E) = k$, so

$$\det(EB) = \det(E)\det(B)$$

Cases 2 and 3 The proofs of the cases where E results from interchanging two rows of I_n or from adding a multiple of one row to another follow the same pattern as Case 1 and are left as exercises. ■

Remark It follows by repeated applications of Lemma 2.3.2 that if B is an $n \times n$ matrix and $E_1, E_2, \ldots, E_r$ are $n \times n$ elementary matrices, then

$$\det(E_1 E_2 \cdots E_r B) = \det(E_1)\det(E_2)\cdots\det(E_r)\det(B) \tag{3}$$

Determinant Test for Invertibility

Our next theorem provides an important criterion for determining whether a matrix is invertible. It also takes us a step closer to establishing Formula (2).

Theorem 2.3.3

A square matrix A is invertible if and only if $\det(A) \neq 0$.

Proof Let R be the reduced row echelon form of A. As a preliminary step, we will show that $\det(A)$ and $\det(R)$ are both zero or both nonzero: Let $E_1, E_2, \ldots, E_r$ be the elementary matrices that correspond to the elementary row operations that produce R from A. Thus

$$R = E_r \cdots E_2 E_1 A$$

and from (3),

$$\det(R) = \det(E_r)\cdots\det(E_2)\det(E_1)\det(A) \tag{4}$$

We pointed out in the margin note that accompanies Theorem 2.2.4 that the determinant of an elementary matrix is nonzero. Thus, it follows from Formula (4) that $\det(A)$ and $\det(R)$ are either both zero or both nonzero, which sets the stage for the main part of the proof. If we assume first that A is invertible, then it follows from Theorem 1.6.4 that $R = I$ and hence that $\det(R) = 1\,(\neq 0)$. This, in turn, implies that $\det(A) \neq 0$, which is what we wanted to show.

Conversely, assume that $\det(A) \neq 0$. It follows from this that $\det(R) \neq 0$, which tells us that R cannot have a row of zeros. Thus, it follows from Theorem 1.4.3 that $R = I$ and hence that A is invertible by Theorem 1.6.4. ■

It follows from Theorems 2.3.3 and 2.2.5 that a square matrix with two proportional rows or two proportional columns is not invertible.

EXAMPLE 3 | Determinant Test for Invertibility

Since the first and third rows of

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

are proportional, $\det(A) = 0$. Thus A is not invertible.

We are now ready for the main result concerning products of matrices.

Theorem 2.3.4

If A and B are square matrices of the same size, then

$$\det(AB) = \det(A)\det(B)$$

Proof We divide the proof into two cases that depend on whether or not A is invertible. If the matrix A is not invertible, then by Theorem 1.6.5 neither is the product AB.

Thus, from Theorem 2.3.3, we have $\det(AB) = 0$ and $\det(A) = 0$, so it follows that $\det(AB) = \det(A)\det(B)$.

Now assume that A is invertible. By Theorem 1.6.4, the matrix A is expressible as a product of elementary matrices, say

$$A = E_1 E_2 \cdots E_r \tag{5}$$

so

$$AB = E_1 E_2 \cdots E_r B$$

Applying (3) to this equation yields

$$\det(AB) = \det(E_1)\det(E_2)\cdots\det(E_r)\det(B)$$

and applying (3) again yields

$$\det(AB) = \det(E_1 E_2 \cdots E_r)\det(B)$$

which, from (5), can be written as $\det(AB) = \det(A)\det(B)$. ■

EXAMPLE 4 | Verifying that $\det(AB) = \det(A)\det(B)$

Consider the matrices

$$A = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 3 \\ 5 & 8 \end{bmatrix}, \quad AB = \begin{bmatrix} 2 & 17 \\ 3 & 14 \end{bmatrix}$$

We leave it for you to verify that

$$\det(A) = 1, \quad \det(B) = -23, \quad \text{and} \quad \det(AB) = -23$$

Thus $\det(AB) = \det(A)\det(B)$, as guaranteed by Theorem 2.3.4.

The following theorem gives a useful relationship between the determinant of an invertible matrix and the determinant of its inverse.

Theorem 2.3.5

If A is invertible, then

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

Historical Note

In 1815 the great French mathematician Augustin Cauchy published a landmark paper in which he gave the first systematic and modern treatment of determinants. It was in that paper that Theorem 2.3.4 was stated and proved in full generality for the first time. Special cases of the theorem had been stated and proved earlier, but it was Cauchy who made the final jump.

[Image: © Bettmann/CORBIS]

Augustin Louis Cauchy (1789–1857)

Proof Since $A^{-1}A = I$, it follows that $\det(A^{-1}A) = \det(I)$. Therefore, we must have $\det(A^{-1})\det(A) = 1$. Since $\det(A) \neq 0$, the proof can be completed by dividing through by $\det(A)$. ■

Adjoint of a Matrix

In a cofactor expansion we compute $\det(A)$ by multiplying the entries in a row or column by their cofactors and adding the resulting products. It turns out that if one multiplies the entries in any row by the corresponding cofactors from a *different* row, the sum of these products is always zero. (This result also holds for columns.) Although we omit the general proof, the next example illustrates this fact.

EXAMPLE 5 | Entries and Cofactors from Different Rows

Let

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

We leave it for you to verify that the cofactors of A are

$$\begin{array}{lll} C_{11} = 12 & C_{12} = 6 & C_{13} = -16 \\ C_{21} = \ \ 4 & C_{22} = 2 & C_{23} = 16 \\ C_{31} = 12 & C_{32} = -10 & C_{33} = 16 \end{array}$$

so, for example, the cofactor expansion of $\det(A)$ along the first row is

$$\det(A) = 3C_{11} + 2C_{12} + (-1)C_{13} = 36 + 12 + 16 = 64$$

and along the first column is

$$\det(A) = 3C_{11} + C_{21} + 2C_{31} = 36 + 4 + 24 = 64$$

Suppose, however, we multiply the entries in the first row by the corresponding cofactors from the *second row* and add the resulting products. The result is

$$3C_{21} + 2C_{22} + (-1)C_{23} = 12 + 4 - 16 = 0$$

Or suppose we multiply the entries in the first column by the corresponding cofactors from the *second column* and add the resulting products. The result is again zero since

$$3C_{12} + 1C_{22} + 2C_{32} = 18 + 2 - 20 = 0$$

Definition 1

If A is any $n \times n$ matrix and C_{ij} is the cofactor of a_{ij}, then the matrix

$$\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}$$

is called the ***matrix of cofactors from A***. The transpose of this matrix is called the ***adjoint of A*** and is denoted by $\text{adj}(A)$.

Historical Note

The use of the term *adjoint* for the transpose of the matrix of cofactors appears to have been introduced by the American mathematician L. E. Dickson in a research paper that he published in 1902.

[Image: Courtesy of the American Mathematical Society (www.ams.org)]

Leonard Eugene Dickson (1874–1954)

EXAMPLE 6 | Adjoint of a 3×3 Matrix

Let

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

As noted in Example 5, the cofactors of A are

$$\begin{array}{lll} C_{11} = 12 & C_{12} = 6 & C_{13} = -16 \\ C_{21} = 4 & C_{22} = 2 & C_{23} = 16 \\ C_{31} = 12 & C_{32} = -10 & C_{33} = 16 \end{array}$$

so the matrix of cofactors is

$$\begin{bmatrix} 12 & 6 & -16 \\ 4 & 2 & 16 \\ 12 & -10 & 16 \end{bmatrix}$$

and the adjoint of A is

$$\operatorname{adj}(A) = \begin{bmatrix} 12 & 4 & 12 \\ 6 & 2 & -10 \\ -16 & 16 & 16 \end{bmatrix}$$

In Theorem 1.4.5 we gave a formula for the inverse of a 2×2 invertible matrix. Our next theorem extends that result to $n \times n$ invertible matrices.

Theorem 2.3.6

Inverse of a Matrix Using Its Adjoint

If A is an invertible matrix, then

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A) \tag{6}$$

Proof We show first that

$$A \operatorname{adj}(A) = \det(A)I$$

Consider the product

$$A \operatorname{adj}(A) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{j1} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{j2} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{jn} & \cdots & C_{nn} \end{bmatrix}$$

The entry in the ith row and jth column of the product $A \operatorname{adj}(A)$ is

$$a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn} \tag{7}$$

(see the shaded lines above).

If $i = j$, then (7) is the cofactor expansion of $\det(A)$ along the ith row of A (Theorem 2.1.1), and if $i \neq j$, then the a's and the cofactors come from different rows of A, so the value of (7) is zero (as illustrated in Example 5). Therefore,

$$A \operatorname{adj}(A) = \begin{bmatrix} \det(A) & 0 & \cdots & 0 \\ 0 & \det(A) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \det(A) \end{bmatrix} = \det(A)I \tag{8}$$

Since A is invertible, $\det(A) \neq 0$. Therefore, Equation (8) can be rewritten as

$$\frac{1}{\det(A)}[A \operatorname{adj}(A)] = I \quad \text{or} \quad A\left[\frac{1}{\det(A)}\operatorname{adj}(A)\right] = I$$

Multiplying both sides on the left by A^{-1} yields

$$A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A) \quad \blacksquare$$

EXAMPLE 7 | Using the Adjoint to Find an Inverse Matrix

Use Formula (6) to find the inverse of the matrix A in Example 6.

Solution We showed in Example 5 that $\det(A) = 64$. Thus,

$$A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A) = \frac{1}{64}\begin{bmatrix} 12 & 4 & 12 \\ 6 & 2 & -10 \\ -16 & 16 & 16 \end{bmatrix} = \begin{bmatrix} \frac{12}{64} & \frac{4}{64} & \frac{12}{64} \\ \frac{6}{64} & \frac{2}{64} & -\frac{10}{64} \\ -\frac{16}{64} & \frac{16}{64} & \frac{16}{64} \end{bmatrix}$$

Cramer's Rule

Our next theorem uses the formula for the inverse of an invertible matrix to produce a formula, called ***Cramer's rule***, for the solution of a linear system $A\mathbf{x} = \mathbf{b}$ of n equations in n unknowns in the case where the coefficient matrix A is invertible (or, equivalently, when $\det(A) \neq 0$).

Theorem 2.3.7

Cramer's Rule

If $A\mathbf{x} = \mathbf{b}$ is a system of n linear equations in n unknowns such that $\det(A) \neq 0$, then the system has a unique solution. This solution is

$$x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \ldots, \quad x_n = \frac{\det(A_n)}{\det(A)}$$

where A_j is the matrix obtained by replacing the entries in the jth column of A by the entries in the matrix

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Proof If $\det(A) \neq 0$, then A is invertible, and by Theorem 1.6.2, $\mathbf{x} = A^{-1}\mathbf{b}$ is the unique solution of $A\mathbf{x} = \mathbf{b}$. Therefore, by Theorem 2.3.6 we have

$$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{\det(A)}\operatorname{adj}(A)\mathbf{b} = \frac{1}{\det(A)}\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Multiplying the matrices out gives

$$\mathbf{x} = \frac{1}{\det(A)}\begin{bmatrix} b_1C_{11} + b_2C_{21} + \cdots + b_nC_{n1} \\ b_1C_{12} + b_2C_{22} + \cdots + b_nC_{n2} \\ \vdots \quad\quad \vdots \quad\quad\quad \vdots \\ b_1C_{1n} + b_2C_{2n} + \cdots + b_nC_{nn} \end{bmatrix}$$

The entry in the jth row of $\mathbf{x}$ is therefore

$$x_j = \frac{b_1C_{1j} + b_2C_{2j} + \cdots + b_nC_{nj}}{\det(A)} \tag{9}$$

Now let

$$A_j = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j-1} & b_1 & a_{1j+1} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j-1} & b_2 & a_{2j+1} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj-1} & b_n & a_{nj+1} & \cdots & a_{nn} \end{bmatrix}$$

Historical Note

Variations of Cramer's rule were fairly well known before the Swiss mathematician discussed it in work he published in 1750. It was Cramer's superior notation that popularized the method and led mathematicians to attach his name to it.

[Image: Science Source/Photo Researchers]

Gabriel Cramer
(1704–1752)

Since A_j differs from A only in the jth column, it follows that the cofactors of entries $b_1, b_2, \ldots, b_n$ in A_j are the same as the cofactors of the corresponding entries in the jth column of A. The cofactor expansion of $\det(A_j)$ along the jth column is therefore

$$\det(A_j) = b_1C_{1j} + b_2C_{2j} + \cdots + b_nC_{nj}$$

Substituting this result in (9) gives

$$x_j = \frac{\det(A_j)}{\det(A)} \quad \blacksquare$$

EXAMPLE 8 | Using Cramer's Rule to Solve a Linear System

Use Cramer's rule to solve

$$\begin{aligned} x_1 + + 2x_3 &= 6 \\ -3x_1 + 4x_2 + 6x_3 &= 30 \\ -x_1 - 2x_2 + 3x_3 &= 8 \end{aligned}$$

Solution

$$A = \begin{bmatrix} 1 & 0 & 2 \\ -3 & 4 & 6 \\ -1 & -2 & 3 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 6 & 0 & 2 \\ 30 & 4 & 6 \\ 8 & -2 & 3 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 1 & 6 & 2 \\ -3 & 30 & 6 \\ -1 & 8 & 3 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 1 & 0 & 6 \\ -3 & 4 & 30 \\ -1 & -2 & 8 \end{bmatrix}$$

Therefore,

$$x_1 = \frac{\det(A_1)}{\det(A)} = \frac{-40}{44} = \frac{-10}{11}, \quad x_2 = \frac{\det(A_2)}{\det(A)} = \frac{72}{44} = \frac{18}{11},$$

$$x_3 = \frac{\det(A_3)}{\det(A)} = \frac{152}{44} = \frac{38}{11}$$

For $n > 3$, it is usually more efficient to solve a linear system with n equations in n unknowns by Gauss–Jordan elimination than by Cramer's rule. Its main use is for obtaining properties of solutions of a linear system without actually solving the system.

Equivalence Theorem

In Theorem 1.6.4 we listed five results that are equivalent to the invertibility of a matrix A. We conclude this section by merging Theorem 2.3.3 with that list to produce the following theorem that relates all of the major topics we have studied thus far.

Theorem 2.3.8

Equivalent Statements

If A is an $n \times n$ matrix, then the following statements are equivalent.

(*a*) A is invertible.

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

(*c*) The reduced row echelon form of A is I_n.

(*d*) A can be expressed as a product of elementary matrices.

(*e*) $A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.

(*f*) $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.

(*g*) $\det(A) \neq 0$.

OPTIONAL: We now have all of the machinery necessary to prove the following two results, which we stated without proof in Theorem 1.7.1:

- **Theorem 1.7.1(*c*)** A triangular matrix is invertible if and only if its diagonal entries are all nonzero.
- **Theorem 1.7.1(*d*)** The inverse of an invertible lower triangular matrix is lower triangular, and the inverse of an invertible upper triangular matrix is upper triangular.

Proof of Theorem 1.7.1(c) Let $A = [a_{ij}]$ be a triangular matrix, so that its diagonal entries are

$$a_{11}, a_{22}, \ldots, a_{nn}$$

From Theorem 2.1.2, the matrix A is invertible if and only if

$$\det(A) = a_{11}a_{22}\cdots a_{nn}$$

is nonzero, which is true if and only if the diagonal entries are all nonzero.

Proof of Theorem 1.7.1(d) We will prove the result for upper triangular matrices and leave the lower triangular case for you. Assume that A is upper triangular and invertible. Since

$$A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A)$$

we can prove that A^{-1} is upper triangular by showing that $\operatorname{adj}(A)$ is upper triangular or, equivalently, that the matrix of cofactors is lower triangular. We can do this by showing that every cofactor C_{ij} with $i < j$ (i.e., above the main diagonal) is zero. Since

$$C_{ij} = (-1)^{i+j}M_{ij}$$

it suffices to show that each minor M_{ij} with $i < j$ is zero. For this purpose, let B_{ij} be the matrix that results when the ith row and jth column of A are deleted, so

$$M_{ij} = \det(B_{ij}) \tag{10}$$

From the assumption that $i < j$, it follows that B_{ij} is upper triangular (see Figure 1.7.1). Since A is upper triangular, its $(i+1)$-st row begins with at least i zeros. But the ith row of B_{ij} is the $(i+1)$-st row of A with the entry in the jth column removed. Since $i < j$, none of the first i zeros is removed by deleting the jth column; thus the ith row of B_{ij} starts with at least i zeros, which implies that this row has a zero on the main diagonal. It now follows from Theorem 2.1.2 that $\det(B_{ij}) = 0$ and from (10) that $M_{ij} = 0$. ■

Exercise Set 2.3

In Exercises 1–4, verify that $\det(kA) = k^n \det(A)$.

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

2. $A = \begin{bmatrix} 2 & 2 \\ 5 & -2 \end{bmatrix}$; $k = -4$

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

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

In Exercises 5–6, verify that $\det(AB) = \det(BA)$ *and determine whether the equality* $\det(A+B) = \det(A) + \det(B)$ *holds.*

5. $A = \begin{bmatrix} 2 & 1 & 0 \\ 3 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -1 & 3 \\ 7 & 1 & 2 \\ 5 & 0 & 1 \end{bmatrix}$

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

In Exercises 7–14, use determinants to decide whether the given matrix is invertible.

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

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

9. $A = \begin{bmatrix} 2 & -3 & 5 \\ 0 & 1 & -3 \\ 0 & 0 & 2 \end{bmatrix}$ **10.** $A = \begin{bmatrix} -3 & 0 & 1 \\ 5 & 0 & 6 \\ 8 & 0 & 3 \end{bmatrix}$

11. $A = \begin{bmatrix} 4 & 2 & 8 \\ -2 & 1 & -4 \\ 3 & 1 & 6 \end{bmatrix}$ **12.** $A = \begin{bmatrix} 1 & 0 & -1 \\ 9 & -1 & 4 \\ 8 & 9 & -1 \end{bmatrix}$

13. $A = \begin{bmatrix} 2 & 0 & 0 \\ 8 & 1 & 0 \\ -5 & 3 & 6 \end{bmatrix}$ **14.** $A = \begin{bmatrix} \sqrt{2} & -\sqrt{7} & 0 \\ 3\sqrt{2} & -3\sqrt{7} & 0 \\ 5 & -9 & 0 \end{bmatrix}$

In Exercises 15–18, find the values of k for which the matrix A is invertible.

15. $A = \begin{bmatrix} k-3 & -2 \\ -2 & k-2 \end{bmatrix}$ **16.** $A = \begin{bmatrix} k & 2 \\ 2 & k \end{bmatrix}$

17. $A = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 1 & 6 \\ k & 3 & 2 \end{bmatrix}$ **18.** $A = \begin{bmatrix} 1 & 2 & 0 \\ k & 1 & k \\ 0 & 2 & 1 \end{bmatrix}$

In Exercises 19–23, decide whether the matrix is invertible, and if so, use the adjoint method to find its inverse.

19. $A = \begin{bmatrix} 2 & 5 & 5 \\ -1 & -1 & 0 \\ 2 & 4 & 3 \end{bmatrix}$ **20.** $A = \begin{bmatrix} 2 & 0 & 3 \\ 0 & 3 & 2 \\ -2 & 0 & -4 \end{bmatrix}$

21. $A = \begin{bmatrix} 2 & -3 & 5 \\ 0 & 1 & -3 \\ 0 & 0 & 2 \end{bmatrix}$ **22.** $A = \begin{bmatrix} 2 & 0 & 0 \\ 8 & 1 & 0 \\ -5 & 3 & 6 \end{bmatrix}$

23. $A = \begin{bmatrix} 1 & 3 & 1 & 1 \\ 2 & 5 & 2 & 2 \\ 1 & 3 & 8 & 9 \\ 1 & 3 & 2 & 2 \end{bmatrix}$

In Exercises 24–29, solve by Cramer's rule, where it applies.

24. $\begin{aligned} 7x_1 - 2x_2 &= 3 \\ 3x_1 + x_2 &= 5 \end{aligned}$

25. $\begin{aligned} 4x + 5y \phantom{{}+2z} &= 2 \\ 11x + y + 2z &= 3 \\ x + 5y + 2z &= 1 \end{aligned}$

26. $\begin{aligned} x - 4y + z &= 6 \\ 4x - y + 2z &= -1 \\ 2x + 2y - 3z &= -20 \end{aligned}$

27. $\begin{aligned} x_1 - 3x_2 + x_3 &= 4 \\ 2x_1 - x_2 \phantom{{}-3x_3} &= -2 \\ 4x_1 \phantom{{}-x_2} - 3x_3 &= 0 \end{aligned}$

28. $\begin{aligned} -x_1 - 4x_2 + 2x_3 + x_4 &= -32 \\ 2x_1 - x_2 + 7x_3 + 9x_4 &= 14 \\ -x_1 + x_2 + 3x_3 + x_4 &= 11 \\ x_1 - 2x_2 + x_3 - 4x_4 &= -4 \end{aligned}$

29. $\begin{aligned} 3x_1 - x_2 + x_3 &= 4 \\ -x_1 + 7x_2 - 2x_3 &= 1 \\ 2x_1 + 6x_2 - x_3 &= 5 \end{aligned}$

30. Show that the matrix

$$A = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is invertible for all values of θ; then find A^{-1} using Theorem 2.3.6.

31. Use Cramer's rule to solve for the unknown y without solving for the unknowns x, z, and w.

$$\begin{aligned} 4x + y + z + w &= 6 \\ 3x + 7y - z + w &= 1 \\ 7x + 3y - 5z + 8w &= -3 \\ x + y + z + 2w &= 3 \end{aligned}$$

32. Let $A\mathbf{x} = \mathbf{b}$ be the system in Exercise 31.

a. Solve by Cramer's rule.

b. Solve by Gauss–Jordan elimination.

c. Which method involves fewer computations?

33. Let

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

Assuming that $\det(A) = -7$, find

a. $\det(3A)$ **b.** $\det(A^{-1})$ **c.** $\det(2A^{-1})$

d. $\det((2A)^{-1})$ **e.** $\det\begin{bmatrix} a & g & d \\ b & h & e \\ c & i & f \end{bmatrix}$

34. In each part, find the determinant given that A is a 4×4 matrix for which $\det(A) = -2$.

a. $\det(-A)$ **b.** $\det(A^{-1})$ **c.** $\det(2A^T)$ **d.** $\det(A^3)$

35. In each part, find the determinant given that A is a 3×3 matrix for which $\det(A) = 7$.

a. $\det(3A)$ **b.** $\det(A^{-1})$

c. $\det(2A^{-1})$ **d.** $\det((2A)^{-1})$

Working with Proofs

36. Prove that a square matrix A is invertible if and only if A^TA is invertible.

37. Prove that if A is a square matrix, then

$$\det(A^TA) = \det(AA^T)$$

38. Let $A\mathbf{x} = \mathbf{b}$ be a system of n linear equations in n unknowns with integer coefficients and integer constants. Prove that if $\det(A) = 1$, the solution $\mathbf{x}$ has integer entries.

39. Prove that if $\det(A) = 1$ and all the entries in A are integers, then all the entries in A^{-1} are integers.

True-False Exercises

TF. In parts (a)–(l) determine whether the statement is true or false, and justify your answer.

a. If A is a 3×3 matrix, then $\det(2A) = 2\det(A)$.

b. If A and B are square matrices of the same size such that $\det(A) = \det(B)$, then $\det(A + B) = 2\det(A)$.

c. If A and B are square matrices of the same size and A is invertible, then

$$\det(A^{-1}BA) = \det(B)$$

d. A square matrix A is invertible if and only if $\det(A) = 0$.

e. The matrix of cofactors of A is precisely $[\operatorname{adj}(A)]^T$.

f. For every $n \times n$ matrix A, we have

$$A \cdot \operatorname{adj}(A) = (\det(A))I_n$$

g. If A is a square matrix and the linear system $A\mathbf{x} = \mathbf{0}$ has multiple solutions for $\mathbf{x}$, then $\det(A) = 0$.

h. If A is an $n \times n$ matrix and there exists an $n \times 1$ matrix $\mathbf{b}$ such that the linear system $A\mathbf{x} = \mathbf{b}$ has no solutions, then the reduced row echelon form of A cannot be I_n.

i. If E is an elementary matrix, then $E\mathbf{x} = \mathbf{0}$ has only the trivial solution.

j. If A is an invertible matrix, then the linear system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution if and only if the linear system $A^{-1}\mathbf{x} = \mathbf{0}$ has only the trivial solution.

k. If A is invertible, then $\operatorname{adj}(A)$ must also be invertible.

l. If A has a row of zeros, then so does $\operatorname{adj}(A)$.

Working with Technology

T1. Consider the matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1+\epsilon \end{bmatrix}$$

in which $\epsilon > 0$. Since $\det(A) = \epsilon \neq 0$, it follows from Theorem 2.3.8 that A is invertible. Compute $\det(A)$ for various small nonzero values of ϵ until you find a value that produces $\det(A) = 0$, thereby leading you to conclude erroneously that A is not invertible. Discuss the cause of this.

T2. We know from Exercise 39 that if A is a *square* matrix then $\det(A^TA) = \det(AA^T)$. By experimenting, make a conjecture as to whether this is true if A is not square.

T3. The French mathematician Jacques Hadamard (1865–1963) proved that if A is an $n \times n$ matrix each of whose entries satisfies the condition $|a_{ij}| \le M$, then

$$|\det(A)| \le \sqrt{n^n}\,M^n$$

(***Hadamard's inequality***). For the following matrix A, use this result to find an interval of possible values for $\det(A)$, and then use your technology utility to show that the value of $\det(A)$ falls within this interval.

$$A = \begin{bmatrix} 0.3 & -2.4 & -1.7 & 2.5 \\ 0.2 & -0.3 & -1.2 & 1.4 \\ 2.5 & 2.3 & 0.0 & 1.8 \\ 1.7 & 1.0 & -2.1 & 2.3 \end{bmatrix}$$

Chapter 2 Supplementary Exercises

*In Exercises **1–8**, evaluate the determinant of the given matrix by (a) cofactor expansion and (b) using elementary row operations to introduce zeros into the matrix.*

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

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

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

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

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

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

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

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

9. Evaluate the determinants in Exercises 3–6 by using the arrow technique (see Example 7 in Section 2.1).

10. a. Construct a 4×4 matrix whose determinant is easy to compute using cofactor expansion but hard to evaluate using elementary row operations.

b. Construct a 4×4 matrix whose determinant is easy to compute using elementary row operations but hard to evaluate using cofactor expansion.

11. Use the determinant to decide whether the matrices in Exercises 1–4 are invertible.

12. Use the determinant to decide whether the matrices in Exercises 5–8 are invertible.

*In Exercises **13–15**, find the given determinant by any method.*

13. $\begin{vmatrix} 5 & b-3 \\ b-2 & -3 \end{vmatrix}$

14. $\begin{vmatrix} 3 & -4 & a \\ a^2 & 1 & 2 \\ 2 & a-1 & 4 \end{vmatrix}$

15. $\begin{vmatrix} 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & -4 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 \end{vmatrix}$

16. Solve for x.

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

*In Exercises **17–24**, use the adjoint method (Theorem 2.3.6) to find the inverse of the given matrix, if it exists.*

17. The matrix in Exercise 1.

18. The matrix in Exercise 2.

19. The matrix in Exercise 3.

20. The matrix in Exercise 4.

21. The matrix in Exercise 5.

22. The matrix in Exercise 6.

23. The matrix in Exercise 7.

24. The matrix in Exercise 8.

25. Use Cramer's rule to solve for x' and y' in terms of x and y.

$$\begin{aligned} x &= \tfrac{3}{5}x' - \tfrac{4}{5}y' \\ y &= \tfrac{4}{5}x' + \tfrac{3}{5}y' \end{aligned}$$

26. Use Cramer's rule to solve for x' and y' in terms of x and y.

$$\begin{aligned} x &= x' \cos\theta - y' \sin\theta \\ y &= x' \sin\theta + y' \cos\theta \end{aligned}$$

27. By examining the determinant of the coefficient matrix, show that the following system has a nontrivial solution if and only if $\alpha = \beta$.

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

28. Let A be a 3×3 matrix, each of whose entries is 1 or 0. What is the largest possible value for $\det(A)$?

29. a. For the triangle in the accompanying figure, use trigonometry to show that

$$\begin{aligned} b\cos\gamma + c\cos\beta &= a \\ c\cos\alpha + a\cos\gamma &= b \\ a\cos\beta + b\cos\alpha &= c \end{aligned}$$

and then apply Cramer's rule to show that

$$\cos\alpha = \frac{b^2 + c^2 - a^2}{2bc}$$

b. Use Cramer's rule to obtain similar formulas for $\cos\beta$ and $\cos\gamma$.

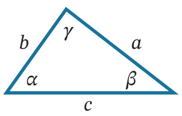


FIGURE Ex-29

30. Use determinants to show that for all real values of λ, the only solution of

$$\begin{aligned} x - 2y &= \lambda x \\ x - y &= \lambda y \end{aligned}$$

is $x = 0, y = 0$.

31. Prove: If A is invertible, then $\operatorname{adj}(A)$ is invertible and

$$[\operatorname{adj}(A)]^{-1} = \frac{1}{\det(A)}A = \operatorname{adj}(A^{-1})$$

32. Prove: If A is an $n \times n$ matrix, then

$$\det[\operatorname{adj}(A)] = [\det(A)]^{n-1}$$

33. Prove: If the entries in each row of an $n \times n$ matrix A add up to zero, then the determinant of A is zero. [*Hint:* Consider the product $A\mathbf{x}$, where $\mathbf{x}$ is the $n \times 1$ matrix, each of whose entries is one.]

34. a. In the accompanying figure, the area of the triangle ABC can be expressed as

$$\text{area } ABC = \text{area } ADEC + \text{area } CEFB - \text{area } ADFB$$

Use this and the fact that the area of a trapezoid equals $\frac{1}{2}$ the altitude times the sum of the parallel sides to show that

$$\text{area } ABC = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$$

[*Note:* In the derivation of this formula, the vertices are labeled such that the triangle is traced counterclockwise proceeding from (x_1, y_1) to (x_2, y_2) to (x_3, y_3). For a clockwise orientation, the determinant above yields the *negative* of the area.]

b. Use the result in (a) to find the area of the triangle with vertices $(3, 3)$, $(4, 0)$, $(-2, -1)$.

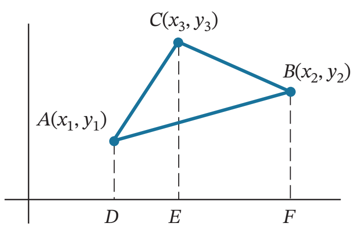


FIGURE Ex-34

35. Use the fact that

$$21375, \quad 38798, \quad 34162, \quad 40223, \quad 79154$$

are all divisible by 19 to show that

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

is divisible by 19 without directly evaluating the determinant.

36. Without directly evaluating the determinant, show that

$$\begin{vmatrix} \sin\alpha & \cos\alpha & \sin(\alpha + \delta) \\ \sin\beta & \cos\beta & \sin(\beta + \delta) \\ \sin\gamma & \cos\gamma & \sin(\gamma + \delta) \end{vmatrix} = 0$$

37. Let $T: R^2 \to R$ be the mapping $(a, b, c, d) \xrightarrow{T} \det\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Is this a linear transformation? Justify your answer.

CHAPTER 3

Euclidean Vector Spaces

CHAPTER CONTENTS

Introduction

Engineers and physicists distinguish between two types of physical quantities—***scalars***, which are quantities that can be described by a numerical value alone, and ***vectors***, which are quantities that require both a number and a direction for their complete physical description. For example, temperature, length, and speed are scalars because they can be fully described by a number that tells "how much"—a temperature of 20°C, a length of 5 cm, or a speed of 75 km/h. In contrast, velocity and force are vectors because they require a number that tells "how much" and a direction that tells "which way"—say, a boat moving at 10 knots in a direction 45° northeast, or a force of 100 lb acting vertically. Although the notions of vectors and scalars that we will study in this text have their origins in physics and engineering, we will be more concerned with using them to build mathematical structures and then applying those structures to such diverse fields as genetics, computer science, economics, telecommunications, and environmental science.

3.1 Vectors in 2-Space, 3-Space, and n-Space

Linear algebra is primarily concerned with two types of mathematical objects, "matrices" and "vectors." In Chapter 1 we discussed the basic properties of matrices, we introduced the idea of viewing n-tuples of real numbers as vectors, and we denoted the set of all such n-tuples as R^n. In this section we will review the basic properties of vectors in two and three dimensions with the goal of extending these properties to vectors in R^n.

Geometric Vectors

Engineers and physicists represent vectors in two dimensions (also called **2-*space***) or in three dimensions (also called **3-*space***) by arrows. The direction of the arrowhead specifies the ***direction*** of the vector and the ***length*** of the arrow specifies the magnitude. Mathematicians call these ***geometric vectors***. The tail of the arrow is called the ***initial point*** of the vector and the tip the ***terminal point*** (**Figure 3.1.1**).

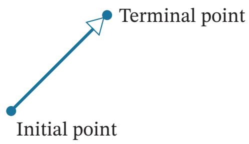


FIGURE 3.1.1

In this text we will denote vectors in boldface type such as **a**, **b**, **v**, **w**, and **x**, and we will denote scalars in lowercase italic type such as a, k, v, w, and x. When we want to indicate that a vector **v** has initial point A and terminal point B, then, as shown in **Figure 3.1.2**, we will write

$$\mathbf{v} = \overrightarrow{AB}$$

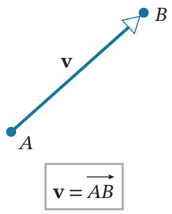
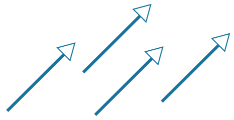


FIGURE 3.1.2

Vectors with the same length and direction, such as those in **Figure 3.1.3**, are said to be ***equivalent***. Since we want a vector to be determined solely by its length and direction, equivalent vectors are regarded as the same vector even though they may be in different, but parallel, positions. Equivalent vectors are also said to be ***equal***, which we indicate by writing

$$\mathbf{v} = \mathbf{w}$$



Equivalent vectors

FIGURE 3.1.3

The vector whose initial and terminal points coincide has length zero, so we call this the ***zero vector*** and denote it by **0**. The zero vector has no natural direction, so we will agree that it can be assigned any direction that is convenient for the problem at hand.

Vector Addition

There are a number of important algebraic operations on vectors, all of which have their origin in laws of physics.

Parallelogram Rule for Vector Addition

If **v** and **w** are vectors in 2-space or 3-space that are positioned so their initial points coincide, then the two vectors form adjacent sides of a parallelogram, and the ***sum*** $\mathbf{v} + \mathbf{w}$ is the vector represented by the arrow from the common initial point of **v** and **w** to the opposite vertex of the parallelogram (**Figure 3.1.4*a***).

Here is another way to form the sum of two vectors.

Triangle Rule for Vector Addition

If **v** and **w** are vectors in 2-space or 3-space that are positioned so the initial point of **w** is at the terminal point of **v**, then the ***sum*** $\mathbf{v} + \mathbf{w}$ is represented by the arrow from the initial point of **v** to the terminal point of **w** (**Figure 3.1.4*b***).

In **Figure 3.1.4*c*** we have constructed the sums $\mathbf{v} + \mathbf{w}$ and $\mathbf{w} + \mathbf{v}$ by the triangle rule. This construction makes it evident that

$$\boxed{\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}} \tag{1}$$

and that the sum obtained by the triangle rule is the same as the sum obtained by the parallelogram rule.

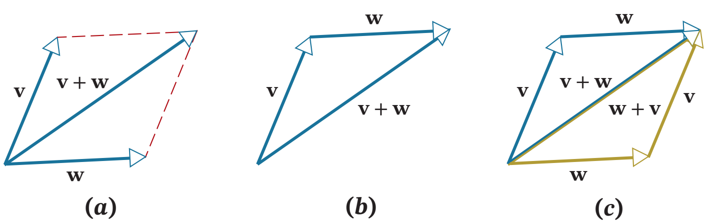


FIGURE 3.1.4

Vector addition can also be viewed as a process of translating points.

Vector Addition Viewed as Translation

If $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{v}+\mathbf{w}$ are positioned so their initial points coincide, then the terminal point of $\mathbf{v}+\mathbf{w}$ can be viewed in two ways:

1. The terminal point of $\mathbf{v}+\mathbf{w}$ is the point that results when the terminal point of $\mathbf{v}$ is translated in the direction of $\mathbf{w}$ by a distance equal to the length of $\mathbf{w}$ (**Figure 3.1.5*a***).
2. The terminal point of $\mathbf{v}+\mathbf{w}$ is the point that results when the terminal point of $\mathbf{w}$ is translated in the direction of $\mathbf{v}$ by a distance equal to the length of $\mathbf{v}$ (**Figure 3.1.5*b***).

Accordingly, we say that the sum $\mathbf{v}+\mathbf{w}$ is the ***translation of*** $\mathbf{v}$ ***by*** $\mathbf{w}$ or, alternatively, the ***translation of*** $\mathbf{w}$ ***by*** $\mathbf{v}$.

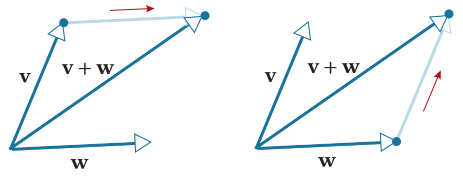


FIGURE 3.1.5

Vector Subtraction

In ordinary arithmetic we can write $a - b = a + (-b)$, which expresses subtraction in terms of addition. There is an analogous idea in vector arithmetic.

Vector Subtraction

The ***negative*** of a vector $\mathbf{v}$, denoted by $-\mathbf{v}$, is the vector that has the same length as $\mathbf{v}$ but is oppositely directed (**Figure 3.1.6*a***), and the ***difference*** of $\mathbf{v}$ from $\mathbf{w}$, denoted by $\mathbf{w}-\mathbf{v}$, is defined to be the sum

$$\mathbf{w}-\mathbf{v} = \mathbf{w} + (-\mathbf{v}) \tag{2}$$

The *difference* of $\mathbf{v}$ from $\mathbf{w}$ can be obtained geometrically by the parallelogram method shown in **Figure 3.1.6*b***, or more directly by positioning $\mathbf{w}$ and $\mathbf{v}$ so their initial points coincide and drawing the vector from the terminal point of $\mathbf{v}$ to the terminal point of $\mathbf{w}$ (**Figure 3.1.6*c***).

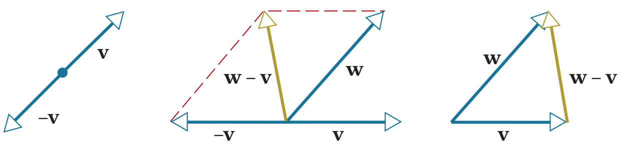


FIGURE 3.1.6

Scalar Multiplication

Sometimes there is a need to change the length of a vector or change its length and reverse its direction. This is accomplished by a type of multiplication in which vectors are multiplied by real numbers, called ***scalars***. As an example, the product $2\mathbf{v}$ denotes the vector that has the same direction as $\mathbf{v}$ but twice the length, and the product $-2\mathbf{v}$ denotes the vector that is oppositely directed to $\mathbf{v}$ and has twice the length. Here is the general result.

> **Scalar Multiplication**
>
> *If* $\mathbf{v}$ is a nonzero vector in 2-space or 3-space, and *if* k is a nonzero scalar, then we define the ***scalar product of*** $\mathbf{v}$ ***by*** $\mathbf{k}$ to be the vector whose length is $|k|$ times the length of $\mathbf{v}$ and whose direction is the same as that of $\mathbf{v}$ if k is positive and opposite to that of $\mathbf{v}$ if k is negative. If $k = 0$ or $\mathbf{v} = \mathbf{0}$, then we define $k\mathbf{v}$ to be $\mathbf{0}$.

Figure 3.1.7 shows the geometric relationship between a vector $\mathbf{v}$ and some of its scalar multiples. In particular, observe that $(-1)\mathbf{v}$ has the same length as $\mathbf{v}$ but is oppositely directed; therefore,

$$\boxed{(-1)\mathbf{v} = -\mathbf{v}} \tag{3}$$

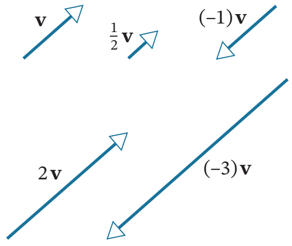


FIGURE 3.1.7

Parallel and Collinear Vectors

Suppose that $\mathbf{v}$ and $\mathbf{w}$ are vectors in 2-space or 3-space with a common initial point. If one of the vectors is a scalar multiple of the other, then the vectors lie on a common line, so it is reasonable to say that they are *collinear* (**Figure 3.1.8*a***). However, if we translate one of the vectors, as indicated in **Figure 3.1.8*b***, then the vectors are *parallel* but no longer collinear. This creates a linguistic problem because translating a vector does not change it. The only way to resolve this problem is to agree that the terms *parallel and collinear mean the same thing when applied to vectors*. Although the vector $\mathbf{0}$ has no clearly defined direction, we will regard it as parallel to all vectors when convenient.

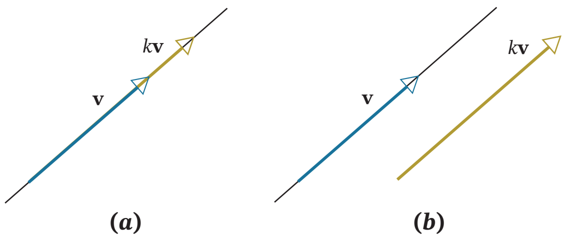


FIGURE 3.1.8

Sums of Three or More Vectors

Vector addition satisfies the ***associative law for addition***, meaning that when we add three vectors, say $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$, it does not matter which two we add first; that is,

$$\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$$

It follows from this that there is no ambiguity in the expression $\mathbf{u} + \mathbf{v} + \mathbf{w}$ because the same result is obtained no matter how the vectors are grouped.

A simple way to construct $\mathbf{u} + \mathbf{v} + \mathbf{w}$ is to place the vectors "tip to tail" in succession and then draw the vector from the initial point of $\mathbf{u}$ to the terminal point of $\mathbf{w}$ (**Figure 3.1.9*a***). The tip-to-tail method also works for four or more vectors (**Figure 3.1.9*b***). The tip-to-tail method makes it evident that if $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in 3-space with a *common initial point*, then $\mathbf{u} + \mathbf{v} + \mathbf{w}$ is the diagonal of the parallelepiped that has the three vectors as adjacent sides (**Figure 3.1.9*c***).

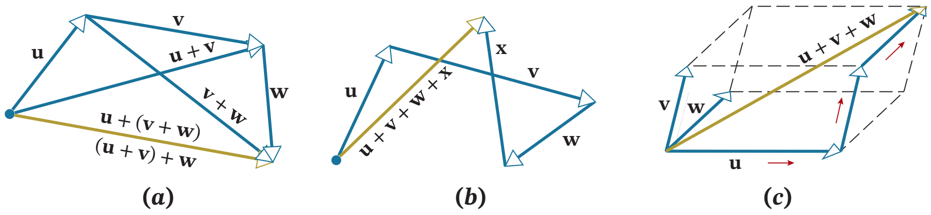


FIGURE 3.1.9

Vectors in Coordinate Systems

Up until now we have discussed vectors without reference to a coordinate system. However, as we will soon see, computations with vectors are much simpler to perform if a coordinate system is present to work with.

If a vector **v** in 2-space or 3-space is positioned with its initial point at the origin of a rectangular coordinate system, then the vector is completely determined by the coordinates of its terminal point (**Figure 3.1.10**). We call these coordinates the ***components*** of **v** relative to the coordinate system. We will write $\mathbf{v} = (v_1, v_2)$ to denote a vector **v** in 2-space with components (v_1, v_2) and $\mathbf{v} = (v_1, v_2, v_3)$ to denote a vector **v** in 3-space with components (v_1, v_2, v_3).

The component forms of the zero vector are $\mathbf{0} = (0, 0)$ in 2-space and $\mathbf{0} = (0, 0, 0)$ in 3-space.

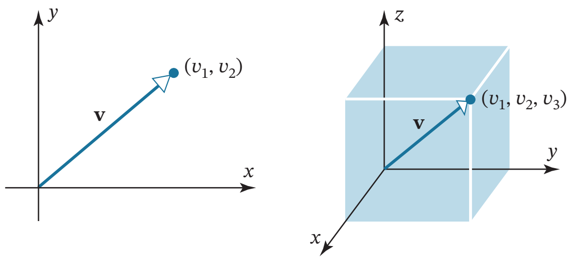


FIGURE 3.1.10

It should be evident geometrically that two vectors in 2-space or 3-space are equivalent if and only if they have the same terminal point when their initial points are at the origin. Algebraically, this means that two vectors are equivalent if and only if their corresponding components are equal. Thus, for example, the vectors

$$\mathbf{v} = (v_1, v_2, v_3) \quad \text{and} \quad \mathbf{w} = (w_1, w_2, w_3)$$

in 3-space are equivalent if and only if

$$v_1 = w_1, \quad v_2 = w_2, \quad v_3 = w_3$$

Remark It may have occurred to you that an ordered pair (v_1, v_2) can represent either a vector with *components* v_1 and v_2 or a point with *coordinates* v_1 and v_2 (and similarly for ordered triples). Both are valid geometric interpretations, so the appropriate choice will depend on the geometric viewpoint that we want to emphasize (**Figure 3.1.11**).

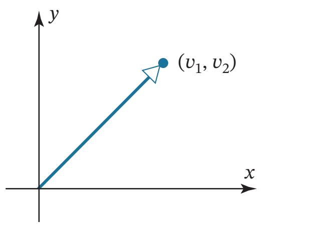


FIGURE 3.1.11 The ordered pair (v_1, v_2) can represent a point or a vector.

Vectors Whose Initial Point Is Not at the Origin

It is sometimes necessary to consider vectors whose initial points are not at the origin. If $\overrightarrow{P_1P_2}$ denotes the vector with initial point $P_1(x_1, y_1)$ and terminal point $P_2(x_2, y_2)$, then the components of this vector are given by the formula

$$\boxed{\overrightarrow{P_1P_2} = (x_2 - x_1, y_2 - y_1)} \tag{4}$$

That is, the components of $\overrightarrow{P_1P_2}$ are obtained by subtracting the coordinates of the initial point from the coordinates of the terminal point. For example, in **Figure 3.1.12** the vector $\overrightarrow{P_1P_2}$ is the difference of vectors $\overrightarrow{OP_2}$ and $\overrightarrow{OP_1}$, so

$$\overrightarrow{P_1P_2} = \overrightarrow{OP_2} - \overrightarrow{OP_1} = (x_2, y_2) - (x_1, y_1) = (x_2 - x_1, y_2 - y_1)$$

As you might expect, the components of a vector in 3-space that has initial point $P_1(x_1, y_1, z_1)$ and terminal point $P_2(x_2, y_2, z_2)$ are given by

$$\boxed{\overrightarrow{P_1P_2} = (x_2 - x_1, y_2 - y_1, z_2 - z_1)} \tag{5}$$

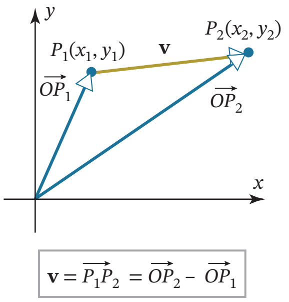


FIGURE 3.1.12

EXAMPLE 1 | Finding the Components of a Vector

The components of the vector $\mathbf{v} = \overrightarrow{P_1P_2}$ with initial point $P_1(2, -1, 4)$ and terminal point $P_2(7, 5, -8)$ are

$$\mathbf{v} = (7 - 2, 5 - (-1), (-8) - 4) = (5, 6, -12)$$

n-Space

The idea of using ordered pairs and triples of real numbers to represent points in two-dimensional space and three-dimensional space was well known in the eighteenth and nineteenth centuries. By the dawn of the twentieth century, mathematicians and physicists were exploring the use of "higher dimensional" spaces in mathematics and physics. Today, even the layman is familiar with the notion of time as a fourth dimension, an idea used by Albert Einstein in developing the general theory of relativity. Today, physicists working in the field of "string theory" commonly use 11-dimensional space in their quest for a unified theory that will explain how the fundamental forces of nature work. Much of the remaining work in this section is concerned with extending the notion of space to n dimensions.

To explore these ideas further, we start with some terminology and notation. The set of all real numbers can be viewed geometrically as points on a line. It is called the ***real line*** and is denoted by R or R^1. The superscript reinforces the intuitive idea that a line is one-dimensional. The set of all ordered pairs of real numbers (called **2-*tuples***) and the set of all ordered triples of real numbers (called **3-*tuples***) are denoted by R^2 and R^3, respectively. The superscript reinforces the idea that the ordered pairs correspond to points in the plane (two-dimensional) and ordered triples to points in space (three-dimensional). The following definition extends this idea.

Definition 1

If n is a positive integer, then an ***ordered n-tuple*** is a sequence of n real numbers $(v_1, v_2, \ldots, v_n)$. The set of all ordered n-tuples is called ***real n-space*** and is denoted by R^n.

Remark You can think of the numbers in an n-tuple $(v_1, v_2, \ldots, v_n)$ as either the coordinates of a *generalized point* or the components of a *generalized vector*, depending on the geometric image you want to bring to mind—the choice makes no difference mathematically, since it is the algebraic properties of n-tuples that are of concern.

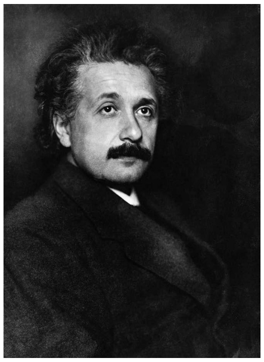
Here are some typical applications that lead to n-tuples.

- **Experimental Data**—A scientist performs an experiment and makes n numerical measurements each time the experiment is performed. The result of each experiment can be regarded as a vector $\mathbf{y} = (y_1, y_2, \dots, y_n)$ in R^n in which $y_1, y_2, \dots, y_n$ are the measured values.
- **Storage and Warehousing**—A national trucking company has 15 depots for storing and servicing its trucks. At each point in time the distribution of trucks in the service depots can be described by a 15-tuple $\mathbf{x} = (x_1, x_2, \dots, x_{15})$ in which x_1 is the number of trucks in the first depot, x_2 is the number in the second depot, and so forth.
- **Electrical Circuits**—A certain kind of processing chip is designed to receive four input voltages and produce three output voltages in response. The input voltages can be regarded as vectors in R^4 and the output voltages as vectors in R^3. Thus, the chip can be viewed as a device that transforms an input vector $\mathbf{v} = (v_1, v_2, v_3, v_4)$ in R^4 into an output vector $\mathbf{w} = (w_1, w_2, w_3)$ in R^3.
- **Graphical Images**—One way in which color images are created on computer screens is by assigning each pixel (an addressable point on the screen) three numbers that describe the ***hue***, ***saturation***, and ***brightness*** of the pixel. Thus, a complete color image can be viewed as a set of 5-tuples of the form $\mathbf{v} = (x, y, h, s, b)$ in which x and y are the screen coordinates of a pixel and h, s, and b are its hue, saturation, and brightness.
- **Economics**—One approach to economic analysis is to divide an economy into sectors (manufacturing, services, utilities, and so forth) and measure the output of each sector by a dollar value. Thus, in an economy with 10 sectors the economic output of the entire economy can be represented by a 10-tuple $\mathbf{s} = (s_1, s_2, \dots, s_{10})$ in which the numbers $s_1, s_2, \dots, s_{10}$ are the outputs of the individual sectors.
- **Mechanical Systems**—Suppose that six particles move along the same coordinate line so that their coordinates are $x_1, x_2, \dots, x_6$ and their velocities are $v_1, v_2, \dots, v_6$, respectively at time t. This information can be represented by the vector

$$\mathbf{v} = (x_1, x_2, x_3, x_4, x_5, x_6, v_1, v_2, v_3, v_4, v_5, v_6, t)$$

in R^{13}. This vector is called the ***state*** of the particle system at time t.

Historical Note



Albert Einstein (1879–1955)

The German-born physicist Albert Einstein immigrated to the United States in 1935, where he settled at Princeton University. Einstein spent the last three decades of his life working unsuccessfully at producing a *unified field theory* that would establish an underlying link between the forces of gravity and electromagnetism. Recently, physicists have made progress on the problem using a framework known as *string theory*. In this theory the smallest, indivisible components of the universe are not particles but loops that behave like vibrating strings. Whereas Einstein's space-time universe was four-dimensional, strings reside in an 11-dimensional world that is the focus of current research.

Operations on Vectors in R^n

Our next goal is to define useful operations on vectors in R^n. These operations will all be natural extensions of the familiar operations on vectors in R^2 and R^3. We will denote a vector $\mathbf{v}$ in R^n using the notation

$$\mathbf{v} = (v_1, v_2, \ldots, v_n)$$

and we will call $\mathbf{0} = (0, 0, \ldots, 0)$ the ***zero vector***.

We noted earlier that in R^2 and R^3 two vectors are equivalent (equal) if and only if their corresponding components are the same. Thus, we make the following definition.

Definition 2

Vectors $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ and $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ in R^n are said to be ***equal*** (also called ***equivalent***) if

$$v_1 = w_1, \quad v_2 = w_2, \ldots, \quad v_n = w_n$$

We indicate this by writing $\mathbf{v} = \mathbf{w}$.

EXAMPLE 2 | Equality of Vectors

The vectors

$$\mathbf{v} = (a, b, c, d) \quad \text{and} \quad \mathbf{w} = (1, -4, 2, 7)$$

are equal if and only if $a = 1, b = -4, c = 2$, and $d = 7$.

Our next objective is to define the operations of addition, subtraction, and scalar multiplication for vectors in R^n. To motivate these ideas, we will consider how these operations can be performed on vectors in R^2 using components. By studying **Figure 3.1.13** you should be able to deduce that if $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$, then

$$\mathbf{v} + \mathbf{w} = (v_1 + w_1, v_2 + w_2) \tag{6}$$

$$k\mathbf{v} = (kv_1, kv_2) \tag{7}$$

In particular, it follows from (7) that

$$-\mathbf{v} = (-1)\mathbf{v} = (-v_1, -v_2) \tag{8}$$

and hence that

$$\mathbf{w} - \mathbf{v} = \mathbf{w} + (-\mathbf{v}) = (w_1 - v_1, w_2 - v_2) \tag{9}$$

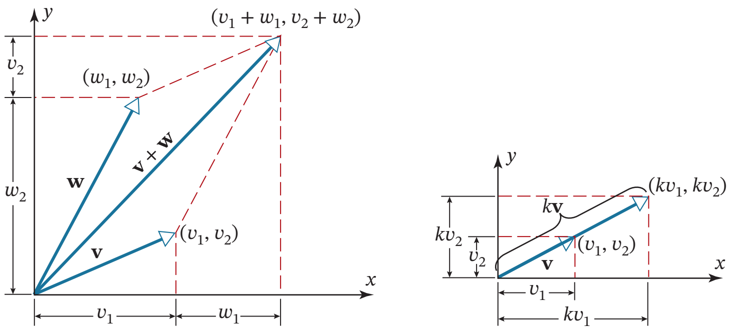


FIGURE 3.1.13

Motivated by Formulas (6)–(9), we make the following definition.

Definition 3

If $\mathbf{v} = (v_1, v_2, \dots, v_n)$ and $\mathbf{w} = (w_1, w_2, \dots, w_n)$ are vectors in R^n, and if k is any scalar, then we define

$$\mathbf{v} + \mathbf{w} = (v_1 + w_1, v_2 + w_2, \dots, v_n + w_n) \tag{10}$$

$$k\mathbf{v} = (kv_1, kv_2, \dots, kv_n) \tag{11}$$

$$-\mathbf{v} = (-v_1, -v_2, \dots, -v_n) \tag{12}$$

$$\mathbf{w} - \mathbf{v} = \mathbf{w} + (-\mathbf{v}) = (w_1 - v_1, w_2 - v_2, \dots, w_n - v_n) \tag{13}$$

In words, vectors are added (or subtracted) by adding (or subtracting) their corresponding components, and a vector is multiplied by a scalar by multiplying each component by that scalar.

EXAMPLE 3 | Algebraic Operations Using Components

If $\mathbf{v} = (1, -3, 2)$ and $\mathbf{w} = (4, 2, 1)$, then

$$\mathbf{v} + \mathbf{w} = (5, -1, 3), \qquad 2\mathbf{v} = (2, -6, 4)$$
$$-\mathbf{w} = (-4, -2, -1), \qquad \mathbf{v} - \mathbf{w} = \mathbf{v} + (-\mathbf{w}) = (-3, -5, 1)$$

The following theorem summarizes the most important properties of vector operations.

Theorem 3.1.1

If $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ are vectors in R^n, and if k and m are scalars, then:

(*a*) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(*b*) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

(*c*) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$

(*d*) $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$

(*e*) $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$

(*f*) $(k + m)\mathbf{u} = k\mathbf{u} + m\mathbf{u}$

(*g*) $k(m\mathbf{u}) = (km)\mathbf{u}$

(*h*) $1\mathbf{u} = \mathbf{u}$

We will prove part (*b*) and leave some of the other proofs as exercises.

Proof (b) Let $\mathbf{u} = (u_1, u_2, \dots, u_n)$, $\mathbf{v} = (v_1, v_2, \dots, v_n)$, and $\mathbf{w} = (w_1, w_2, \dots, w_n)$. Then

$$\begin{aligned}
(\mathbf{u} + \mathbf{v}) + \mathbf{w} &= \big((u_1, u_2, \dots, u_n) + (v_1, v_2, \dots, v_n)\big) + (w_1, w_2, \dots, w_n) \\
&= (u_1 + v_1, u_2 + v_2, \dots, u_n + v_n) + (w_1, w_2, \dots, w_n) && \textbf{[Vector addition]} \\
&= \big((u_1 + v_1) + w_1, (u_2 + v_2) + w_2, \dots, (u_n + v_n) + w_n\big) && \textbf{[Vector addition]} \\
&= \big(u_1 + (v_1 + w_1), u_2 + (v_2 + w_2), \dots, u_n + (v_n + w_n)\big) && \textbf{[Regroup]} \\
&= (u_1, u_2, \dots, u_n) + (v_1 + w_1, v_2 + w_2, \dots, v_n + w_n) && \textbf{[Vector addition]} \\
&= \mathbf{u} + (\mathbf{v} + \mathbf{w}) \quad \blacksquare
\end{aligned}$$

The following additional properties of vectors in R^n can be deduced easily by expressing the vectors in terms of components (verify).

Theorem 3.1.2

If $\mathbf{v}$ is a vector in R^n and k is a scalar, then:

(a) $0\mathbf{v} = \mathbf{0}$

(b) $k\mathbf{0} = \mathbf{0}$

(c) $(-1)\mathbf{v} = -\mathbf{v}$

Calculating Without Components

One of the powerful consequences of Theorems 3.1.1 and 3.1.2 is that they allow calculations to be performed without expressing the vectors in terms of components. For example, suppose that $\mathbf{x}$, $\mathbf{a}$, and $\mathbf{b}$ are vectors in R^n, and we want to solve the vector equation $\mathbf{x} + \mathbf{a} = \mathbf{b}$ for the vector $\mathbf{x}$ without using components. We could proceed as follows:

$$\begin{aligned}
&\mathbf{x} + \mathbf{a} = \mathbf{b} && \textbf{[Given]}\\
&(\mathbf{x} + \mathbf{a}) + (-\mathbf{a}) = \mathbf{b} + (-\mathbf{a}) && \textbf{[Add the negative of a to both sides]}\\
&\mathbf{x} + (\mathbf{a} + (-\mathbf{a})) = \mathbf{b} - \mathbf{a} && \textbf{[Part (}\boldsymbol{b}\textbf{) of Theorem 3.1.1]}\\
&\mathbf{x} + \mathbf{0} = \mathbf{b} - \mathbf{a} && \textbf{[Part (}\boldsymbol{d}\textbf{) of Theorem 3.1.1]}\\
&\mathbf{x} = \mathbf{b} - \mathbf{a} && \textbf{[Part (}\boldsymbol{c}\textbf{) of Theorem 3.1.1]}
\end{aligned}$$

While this method is obviously more cumbersome than computing with components in R^n, it will become important later in the text where we will encounter more general kinds of vectors.

Linear Combinations

Addition, subtraction, and scalar multiplication are frequently used in combination to form new vectors. For example, if $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are vectors in R^n, then the vectors

$$\mathbf{u} = 2\mathbf{v}_1 + 3\mathbf{v}_2 + \mathbf{v}_3 \quad \text{and} \quad \mathbf{w} = 7\mathbf{v}_1 - 6\mathbf{v}_2 + 8\mathbf{v}_3$$

are formed in this way. In general, we make the following definition.

Definition 4

If $\mathbf{w}$ is a vector in R^n, then $\mathbf{w}$ is said to be a ***linear combination*** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$ in R^n if it can be expressed in the form

$$\mathbf{w} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_r\mathbf{v}_r \tag{14}$$

where $k_1, k_2, \dots, k_r$ are scalars. These scalars are called the ***coefficients*** of the linear combination. In the case where $r = 1$, Formula (14) becomes $\mathbf{w} = k_1\mathbf{v}_1$, so that a linear combination of a single vector is just a scalar multiple of that vector.

Note that this definition of a linear combination is consistent with that given in the context of matrices (see Definition 6 in Section 1.3).

Alternative Notations for Vectors

Up to now we have been writing vectors in R^n using the notation

$$\mathbf{v} = (v_1, v_2, \dots, v_n) \tag{15}$$

We call this the ***comma-delimited*** form. However, since a vector in R^n is just a list of its n components in a specific order, any notation that displays those components in the correct order is a valid way of representing the vector. For example, the vector in (15) can be written as

$$\mathbf{v} = [v_1 \quad v_2 \quad \cdots \quad v_n] \tag{16}$$

which is called ***row-vector*** form, or as

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \tag{17}$$

which is called ***column-vector*** form. The choice of notation is often a matter of taste or convenience, but sometimes the nature of a problem will suggest a preferred notation. Notations (15), (16), and (17) will all be used at various places in this text.

Application of Linear Combinations to Color Models

Colors on computer monitors are commonly based on what is called the **RGB *color model***. Colors in this system are created by adding together percentages of the primary colors red (R), green (G), and blue (B). One way to do this is to identify the primary colors with the vectors

$$\begin{aligned} \mathbf{r} &= (1, 0, 0) \quad \text{(pure red)}, \\ \mathbf{g} &= (0, 1, 0) \quad \text{(pure green)}, \\ \mathbf{b} &= (0, 0, 1) \quad \text{(pure blue)} \end{aligned}$$

in R^3 and to create all other colors by forming linear combinations of **r**, **g**, and **b** using coefficients between 0 and 1, inclusive; these coefficients represent the percentage of each pure color in the mix. The set of all such color vectors is called **RGB *space*** or the **RGB *color cube*** (Figure 3.1.14). Thus, each color vector **c** in this cube is expressible as a linear combination of the form

$$\begin{aligned} \mathbf{c} &= k_1\mathbf{r} + k_2\mathbf{g} + k_3\mathbf{b} \\ &= k_1(1, 0, 0) + k_2(0, 1, 0) + k_3(0, 0, 1) \\ &= (k_1, k_2, k_3) \end{aligned}$$

where $0 \le k_i \le 1$. As indicated in the figure, the corners of the cube represent the pure primary colors together with the colors black, white, magenta, cyan, and yellow. The vectors along the diagonal running from black to white correspond to shades of gray.

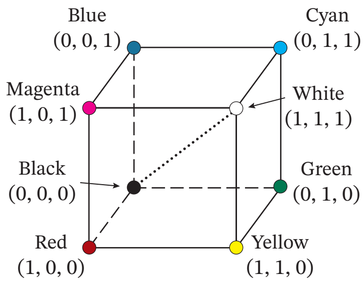
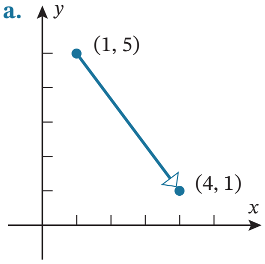


FIGURE 3.1.14

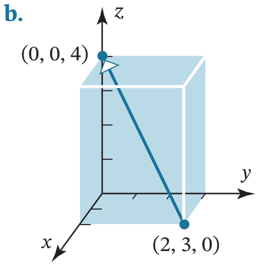
Exercise Set 3.1

In Exercises 1–2, find the components of the vector.

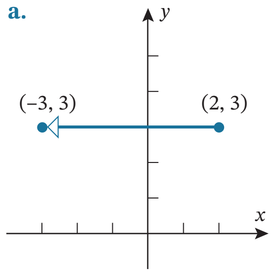
1. a.



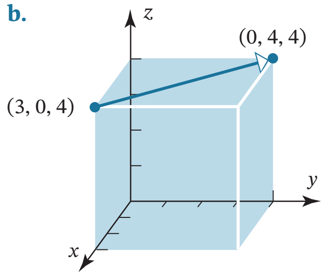
b.



2. a.



b.



In Exercises 3–4, find the components of the vector $\overrightarrow{P_1P_2}$.

3. **a.** $P_1(3,5)$, $P_2(2,8)$ **b.** $P_1(5,-2,1)$, $P_2(2,4,2)$

4. **a.** $P_1(-6,2)$, $P_2(-4,-1)$ **b.** $P_1(0,0,0)$, $P_2(-1,6,1)$

5. **a.** Find the terminal point of the vector that is equivalent to $\mathbf{u}=(1,2)$ and whose initial point is $A(1,1)$.

b. Find the initial point of the vector that is equivalent to $\mathbf{u}=(1,1,3)$ and whose terminal point is $B(-1,-1,2)$.

6. **a.** Find the initial point of the vector that is equivalent to $\mathbf{u}=(1,2)$ and whose terminal point is $B(2,0)$.

b. Find the terminal point of the vector that is equivalent to $\mathbf{u}=(1,1,3)$ and whose initial point is $A(0,2,0)$.

7. Find the initial point P of a nonzero vector $\mathbf{u}=\overrightarrow{PQ}$ with terminal point $Q(3,0,-5)$ and such that

a. $\mathbf{u}$ has the same direction as $\mathbf{v}=(4,-2,-1)$.

b. $\mathbf{u}$ is oppositely directed to $\mathbf{v}=(4,-2,-1)$.

8. Find the terminal point Q of a nonzero vector $\mathbf{u}=\overrightarrow{PQ}$ with initial point $P(-1,3,-5)$ and such that

a. $\mathbf{u}$ has the same direction as $\mathbf{v}=(6,7,-3)$.

b. $\mathbf{u}$ is oppositely directed to $\mathbf{v}=(6,7,-3)$.

9. Let $\mathbf{u}=(4,-1)$, $\mathbf{v}=(0,5)$, and $\mathbf{w}=(-3,-3)$. Find the components of

a. $\mathbf{u}+\mathbf{w}$ **b.** $\mathbf{v}-3\mathbf{u}$

c. $2(\mathbf{u}-5\mathbf{w})$ **d.** $3\mathbf{v}-2(\mathbf{u}+2\mathbf{w})$

10. Let $\mathbf{u}=(-3,1,2)$, $\mathbf{v}=(4,0,-8)$, and $\mathbf{w}=(6,-1,-4)$. Find the components of

a. $\mathbf{v}-\mathbf{w}$ **b.** $6\mathbf{u}+2\mathbf{v}$

c. $-3(\mathbf{v}-8\mathbf{w})$ **d.** $(2\mathbf{u}-7\mathbf{w})-(8\mathbf{v}+\mathbf{u})$

11. Let $\mathbf{u}=(-3,2,1,0)$, $\mathbf{v}=(4,7,-3,2)$, and $\mathbf{w}=(5,-2,8,1)$. Find the components of

a. $\mathbf{v}-\mathbf{w}$ **b.** $-\mathbf{u}+(\mathbf{v}-4\mathbf{w})$

c. $6(\mathbf{u}-3\mathbf{v})$ **d.** $(6\mathbf{v}-\mathbf{w})-(4\mathbf{u}+\mathbf{v})$

12. Let $\mathbf{u}=(1,2,-3,5,0)$, $\mathbf{v}=(0,4,-1,1,2)$, and $\mathbf{w}=(7,1,-4,-2,3)$. Find the components of

a. $\mathbf{v}+\mathbf{w}$ **b.** $3(2\mathbf{u}-\mathbf{v})$

c. $(3\mathbf{u}-\mathbf{v})-(2\mathbf{u}+4\mathbf{w})$ **d.** $\frac{1}{2}(\mathbf{w}-5\mathbf{v}+2\mathbf{u})+\mathbf{v}$

13. Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be the vectors in Exercise 11. Find the components of the vector $\mathbf{x}$ that satisfies the equation $3\mathbf{u}+\mathbf{v}-2\mathbf{w}=3\mathbf{x}+2\mathbf{w}$.

14. Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be the vectors in Exercise 12. Find the components of the vector $\mathbf{x}$ that satisfies the equation $2\mathbf{u}-\mathbf{v}+\mathbf{x}=7\mathbf{x}+\mathbf{w}$.

15. Which of the following vectors in R^6, if any, are parallel to $\mathbf{u}=(-2,1,0,3,5,1)$?

a. $(4,2,0,6,10,2)$

b. $(4,-2,0,-6,-10,-2)$

c. $(0,0,0,0,0,0)$

16. For what value(s) of t, if any, is the given vector parallel to $\mathbf{u}=(4,-1)$?

a. $(8t,-2)$ **b.** $(8t,2t)$ **c.** $(1,t^2)$

17. Let $\mathbf{u}=(1,-1,3,5)$ and $\mathbf{v}=(2,1,0,-3)$. Find scalars a and b so that $a\mathbf{u}+b\mathbf{v}=(1,-4,9,18)$.

18. Let $\mathbf{u}=(2,1,0,1,-1)$ and $\mathbf{v}=(-2,3,1,0,2)$. Find scalars a and b so that $a\mathbf{u}+b\mathbf{v}=(-8,8,3,-1,7)$.

In Exercises 19–20, find scalars c_1, c_2, and c_3 for which the equation is satisfied.

19. $c_1(1,-1,0)+c_2(3,2,1)+c_3(0,1,4)=(-1,1,19)$

20. $c_1(-1,0,2)+c_2(2,2,-2)+c_3(1,-2,1)=(-6,12,4)$

21. Show that there do not exist scalars c_1, c_2, and c_3 such that

$$c_1(-2,9,6)+c_2(-3,2,1)+c_3(1,7,5)=(0,5,4)$$

22. Show that there do not exist scalars c_1, c_2, and c_3 such that

$$c_1(1,0,1,0)+c_2(1,0,-2,1)+c_3(2,0,1,2)=(1,-2,2,3)$$

23. Let P be the point $(2,3,-2)$ and Q the point $(7,-4,1)$.

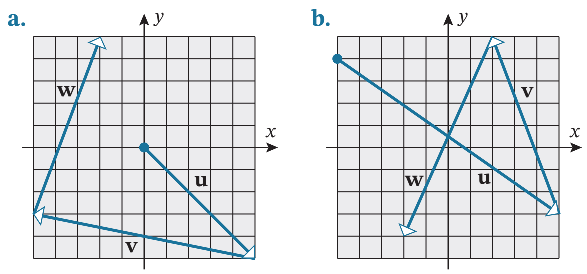
a. Find the midpoint of the line segment connecting the points P and Q.

b. Find the point on the line segment connecting the points P and Q that is $\frac{3}{4}$ of the way from P to Q.

24. In relation to the points P_1 and P_2 in Figure 3.1.12, what can you say about the terminal point of the following vector if its initial point is at the origin?

$$\mathbf{u}=\overrightarrow{OP_1}+\tfrac{1}{2}(\overrightarrow{OP_2}-\overrightarrow{OP_1})$$

25. In each part, find the components of the vector $\mathbf{u}+\mathbf{v}+\mathbf{w}$.



26. Referring to the vectors pictured in Exercise 25, find the components of the vector $\mathbf{u}-\mathbf{v}+\mathbf{w}$.

27. Let P be the point $(1,3,7)$. If the point $(4,0,-6)$ is the midpoint of the line segment connecting P and Q, what is Q?

28. If the sum of three vectors in R^3 is zero, must they lie in the same plane? Explain.

29. Consider the regular hexagon shown in the accompanying figure.

a. What is the sum of the six radial vectors that run from the center to the vertices?

b. How is the sum affected if each radial vector is multiplied by $\frac{1}{2}$?

c. What is the sum of the five radial vectors that remain if **a** is removed?

d. Discuss some variations and generalizations of the result in part (c).

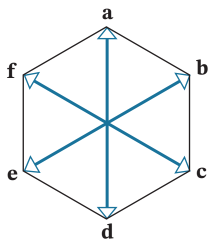


FIGURE Ex-29

30. What is the sum of all radial vectors of a regular n-sided polygon? (See Figure Ex-29.)

Working with Proofs

31. Prove parts (a), (c), and (d) of Theorem 3.1.1.

32. Prove parts (e)–(h) of Theorem 3.1.1.

33. Prove parts (a)–(c) of Theorem 3.1.2.

True-False Exercises

TF. In parts **(a)–(k)** determine whether the statement is true or false, and justify your answer.

a. Two equivalent vectors must have the same initial point.

b. The vectors (a, b) and $(a, b, 0)$ are equivalent.

c. If k is a scalar and $\mathbf{v}$ is a vector, then $\mathbf{v}$ and $k\mathbf{v}$ are parallel if and only if $k \geq 0$.

d. The vectors $\mathbf{v} + (\mathbf{u} + \mathbf{w})$ and $(\mathbf{w} + \mathbf{v}) + \mathbf{u}$ are the same.

e. If $\mathbf{u} + \mathbf{v} = \mathbf{u} + \mathbf{w}$, then $\mathbf{v} = \mathbf{w}$.

f. If a and b are scalars such that $a\mathbf{u} + b\mathbf{v} = \mathbf{0}$, then $\mathbf{u}$ and $\mathbf{v}$ are parallel vectors.

g. Collinear vectors with the same length are equal.

h. If $(a, b, c) + (x, y, z) = (x, y, z)$, then (a, b, c) must be the zero vector.

i. If k and m are scalars and $\mathbf{u}$ and $\mathbf{v}$ are vectors, then

$$(k + m)(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + m\mathbf{v}$$

j. If the vectors $\mathbf{v}$ and $\mathbf{w}$ are given, then the vector equation

$$3(2\mathbf{v} - \mathbf{x}) = 5\mathbf{x} - 4\mathbf{w} + \mathbf{v}$$

can be solved for $\mathbf{x}$.

k. The linear combinations $a_1\mathbf{v}_1 + a_2\mathbf{v}_2$ and $b_1\mathbf{v}_1 + b_2\mathbf{v}_2$ can only be equal if $a_1 = b_1$ and $a_2 = b_2$.

3.2 Norm, Dot Product, and Distance in R^n

In this section we will be concerned with the notions of length and distance as they relate to vectors. We will first discuss these ideas in R^2 and R^3 and then extend them algebraically to R^n.

Norm of a Vector

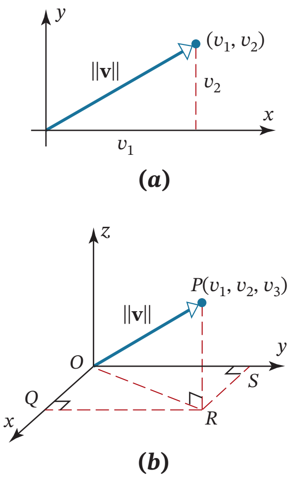


FIGURE 3.2.1

In this text we will denote the length of a vector $\mathbf{v}$ by the symbol $\|\mathbf{v}\|$. As suggested in **Figure 3.2.1*a***, it follows from the Theorem of Pythagoras that the norm of a vector (v_1, v_2) in R^2 is

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2} \tag{1}$$

Similarly, for a vector (v_1, v_2, v_3) in R^3, it follows from **Figure 3.2.1*b*** and two applications of the Theorem of Pythagoras that

$$\|\mathbf{v}\|^2 = (OR)^2 + (RP)^2 = (OQ)^2 + (QR)^2 + (RP)^2 = v_1^2 + v_2^2 + v_3^2$$

and hence that

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2} \tag{2}$$

Motivated by the pattern of Formulas (1) and (2), we make the following definition.

Definition 1

If $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ is a vector in R^n, then the ***norm*** of $\mathbf{v}$ (also called the ***length*** of $\mathbf{v}$ or the ***magnitude*** of $\mathbf{v}$) is denoted by $\|\mathbf{v}\|$, and is defined by the formula

$$\boxed{\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}} \tag{3}$$

EXAMPLE 1 | Calculating Norms

It follows from Formula (2) that the norm of the vector $\mathbf{v} = (-3, 2, 1)$ in R^3 is

$$\|\mathbf{v}\| = \sqrt{(-3)^2 + 2^2 + 1^2} = \sqrt{14}$$

and it follows from Formula (3) that the norm of the vector $\mathbf{v} = (2, -1, 3, -5)$ in R^4 is

$$\|\mathbf{v}\| = \sqrt{2^2 + (-1)^2 + 3^2 + (-5)^2} = \sqrt{39}$$

Our first theorem in this section will generalize to R^n the following three familiar facts about vectors in R^2 and R^3:

- Distances are nonnegative.
- The zero vector is the only vector of length zero.
- Multiplying a vector by a scalar multiplies its length by the absolute value of that scalar.

It is important to recognize that just because these results hold in R^2 and R^3 does not guarantee that they hold in R^n—their validity in R^n must be *proved* using algebraic properties of n-tuples.

Theorem 3.2.1

If $\mathbf{v}$ is a vector in R^n, and if k is any scalar, then:

(a) $\|\mathbf{v}\| \geq 0$

(b) $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$

(c) $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$

We will prove part (c) and leave (a) and (b) as exercises.

Proof (c) If $\mathbf{v} = (v_1, v_2, \ldots, v_n)$, then $k\mathbf{v} = (kv_1, kv_2, \ldots, kv_n)$, so

$$\begin{aligned} \|k\mathbf{v}\| &= \sqrt{(kv_1)^2 + (kv_2)^2 + \cdots + (kv_n)^2} \\ &= \sqrt{(k^2)(v_1^2 + v_2^2 + \cdots + v_n^2)} \\ &= |k|\sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} \\ &= |k|\|\mathbf{v}\| \quad \blacksquare \end{aligned}$$

Unit Vectors

Two nonzero vectors in R^n are said to have the ***same direction*** if each is a positive scalar multiple of the other and ***opposite directions*** if each is a negative scalar multiple of the other. Thus, for example, the vectors $\mathbf{v}_1 = (2, -4, 1, 8)$ and $\mathbf{v}_2 = (1, -2, \frac{1}{2}, 4)$ have the same direction, whereas $\mathbf{w}_1 = (2, -4, 1, 8)$ and $\mathbf{w}_2 = (-1, 2, -\frac{1}{2}, -4)$ have opposite directions.

A vector of norm 1 is called a ***unit vector***. Such vectors are useful for specifying a direction when length is not relevant to the problem at hand. You can obtain a unit vector in a desired direction by choosing any *nonzero* vector $\mathbf{v}$ in that direction and multiplying $\mathbf{v}$ by the reciprocal of its length. For example, if $\mathbf{v}$ is a vector of length 2 in R^2 or R^3, then $\frac{1}{2}\mathbf{v}$ is a unit vector in the same direction as $\mathbf{v}$. More generally, if $\mathbf{v}$ is any nonzero vector in R^n, then

$$\boxed{\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}} \tag{4}$$

Warning Sometimes you will see Formula (4) expressed as

$$\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$$

This is just a more compact way of writing that formula and is *not* intended to convey that $\mathbf{v}$ is being divided by $\|\mathbf{v}\|$.

defines a unit vector that is in the same direction as $\mathbf{v}$. We can confirm that (4) is a unit vector by applying part (*c*) of Theorem 3.2.1 with $k = 1/\|\mathbf{v}\|$ to obtain

$$\|\mathbf{u}\| = \|k\mathbf{v}\| = |k|\|\mathbf{v}\| = k\|\mathbf{v}\| = \frac{1}{\|\mathbf{v}\|}\|\mathbf{v}\| = 1$$

The process of multiplying a nonzero vector by the reciprocal of its length to obtain a unit vector is called ***normalizing*** $\mathbf{v}$.

EXAMPLE 2 | Normalizing a Vector

Find the unit vector $\mathbf{u}$ that has the same direction as $\mathbf{v} = (2, 2, -1)$.

Solution The vector $\mathbf{v}$ has length

$$\|\mathbf{v}\| = \sqrt{2^2 + 2^2 + (-1)^2} = 3$$

Thus, from (4)

$$\mathbf{u} = \tfrac{1}{3}(2, 2, -1) = \left(\tfrac{2}{3}, \tfrac{2}{3}, -\tfrac{1}{3}\right)$$

As a check, you may want to confirm that $\|\mathbf{u}\| = 1$.

The Standard Unit Vectors

When a rectangular coordinate system is introduced in R^2 or R^3, the unit vectors in the positive directions of the coordinate axes are called the ***standard unit vectors***. In R^2 these vectors are denoted by

$$\mathbf{i} = (1, 0) \quad \text{and} \quad \mathbf{j} = (0, 1)$$

and in R^3 by

$$\mathbf{i} = (1, 0, 0), \quad \mathbf{j} = (0, 1, 0), \quad \text{and} \quad \mathbf{k} = (0, 0, 1)$$

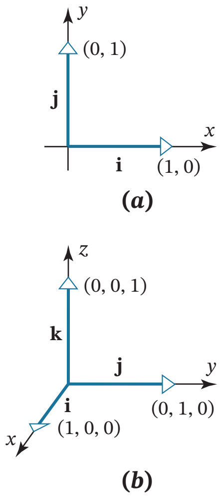


FIGURE 3.2.2

(Figure 3.2.2). Every vector $\mathbf{v} = (v_1, v_2)$ in R^2 and every vector $\mathbf{v} = (v_1, v_2, v_3)$ in R^3 can be expressed as a linear combination of standard unit vectors by writing

$$\mathbf{v} = (v_1, v_2) = v_1(1, 0) + v_2(0, 1) = v_1\mathbf{i} + v_2\mathbf{j} \tag{5}$$

$$\mathbf{v} = (v_1, v_2, v_3) = v_1(1, 0, 0) + v_2(0, 1, 0) + v_3(0, 0, 1) = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k} \tag{6}$$

Moreover, we can generalize these formulas to R^n by defining the ***standard unit vectors in R^n*** to be

$$\mathbf{e}_1 = (1, 0, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, 0, \ldots, 0), \ldots, \quad \mathbf{e}_n = (0, 0, 0, \ldots, 1) \tag{7}$$

in which case every vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ in R^n can be expressed as

$$\boxed{\mathbf{v} = (v_1, v_2, \ldots, v_n) = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n} \tag{8}$$

EXAMPLE 3 | Linear Combinations of Standard Unit Vectors

$$(2, -3, 4) = 2\mathbf{i} - 3\mathbf{j} + 4\mathbf{k}$$
$$(7, 3, -4, 5) = 7\mathbf{e}_1 + 3\mathbf{e}_2 - 4\mathbf{e}_3 + 5\mathbf{e}_4$$

Distance in R^n

If P_1 and P_2 are points in R^2 or R^3, then the length of the vector $\overrightarrow{P_1P_2}$ is equal to the distance d between the two points (**Figure 3.2.3**). Specifically, if $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ are points in R^2, then Formula (4) of Section 3.1 implies that

$$d = \|\overrightarrow{P_1P_2}\| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{9}$$

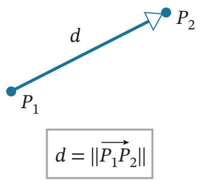


FIGURE 3.2.3

This is the familiar distance formula from analytic geometry. Similarly, the distance between the points $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ in 3-space is

$$\boxed{d(\mathbf{u}, \mathbf{v}) = \|\overrightarrow{P_1P_2}\| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}} \tag{10}$$

Motivated by Formulas (9) and (10), we make the following definition.

Definition 2

If $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ are points in R^n, then we denote the ***distance*** between $\mathbf{u}$ and $\mathbf{v}$ by $d(\mathbf{u}, \mathbf{v})$ and define it to be

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2} \tag{11}$$

We noted in the previous section that n-tuples can be viewed either as vectors or points in R^n. In Definition 2 we chose to describe them as points, as that seemed the more natural interpretation.

EXAMPLE 4 | Calculating Distance in R^n

If

$$\mathbf{u} = (1, 3, -2, 7) \quad \text{and} \quad \mathbf{v} = (0, 7, 2, 2)$$

then the distance between $\mathbf{u}$ and $\mathbf{v}$ is

$$d(\mathbf{u}, \mathbf{v}) = \sqrt{(1-0)^2 + (3-7)^2 + (-2-2)^2 + (7-2)^2} = \sqrt{58}$$

Dot Product

Our next objective is to define a useful multiplication operation on vectors in R^2 and R^3 and then extend that operation to R^n. To do this we will first need to define exactly what we mean by the "angle" between two vectors in R^2 or R^3. For this purpose, let $\mathbf{u}$ and $\mathbf{v}$ be nonzero vectors in R^2 or R^3 that have been positioned so that their initial points coincide. We define the ***angle between* u *and* v** to be the angle θ determined by **u** and **v** that satisfies the inequalities $0 \leq \theta \leq \pi$ (**Figure 3.2.4**).

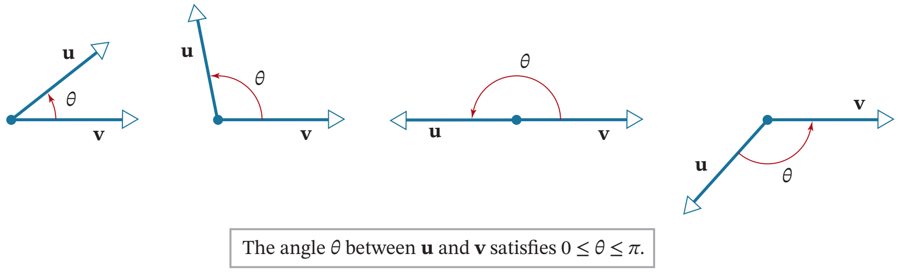


FIGURE 3.2.4

Definition 3

If **u** and **v** are nonzero vectors in R^2 or R^3, and if θ is the angle between **u** and **v**, then the ***dot product*** (also called the ***Euclidean inner product***) of **u** and **v** is denoted by $\mathbf{u} \cdot \mathbf{v}$ and is defined as

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta \tag{12}$$

If $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$, then we define $\mathbf{u} \cdot \mathbf{v}$ to be 0.

If **u** and **v** are nonzero, then the sign of the dot product reveals information about the angle θ that we can obtain by rewriting Formula (12) as

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} \tag{13}$$

Since $0 \leq \theta \leq \pi$, it follows from Formula (13) and properties of the cosine function that

- θ is acute if $\mathbf{u} \cdot \mathbf{v} > 0$.
- θ is obtuse if $\mathbf{u} \cdot \mathbf{v} < 0$.
- $\theta = \pi/2$ if $\mathbf{u} \cdot \mathbf{v} = 0$.

EXAMPLE 5 | Dot Product

Find the dot product of the vectors shown in **Figure 3.2.5**.

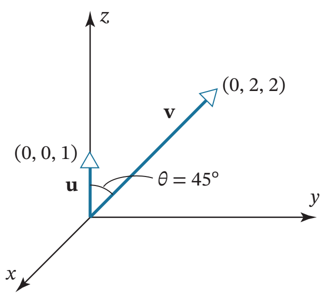


FIGURE 3.2.5

Solution The lengths of the vectors are

$$\|\mathbf{u}\| = 1 \quad \text{and} \quad \|\mathbf{v}\| = \sqrt{8} = 2\sqrt{2}$$

and the cosine of the angle θ between them is

$$\cos(45^\circ) = 1/\sqrt{2}$$

Thus, it follows from Formula (12) that

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta = (1)(2\sqrt{2})(1/\sqrt{2}) = 2$$

Component Form of the Dot Product

For computational purposes it is desirable to have a formula that expresses the dot product of two vectors in terms of components. We will derive such a formula for vectors in 3-space; the derivation for vectors in 2-space is similar.

Let $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ be two nonzero vectors. If, as shown in Figure 3.2.6, θ is the angle between $\mathbf{u}$ and $\mathbf{v}$, then the law of cosines yields

$$\|\overrightarrow{PQ}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta \tag{14}$$

Since $\overrightarrow{PQ} = \mathbf{v} - \mathbf{u}$, we can rewrite (14) as

$$\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta = \tfrac{1}{2}(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{v} - \mathbf{u}\|^2)$$

or

$$\mathbf{u}\cdot\mathbf{v} = \tfrac{1}{2}(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{v} - \mathbf{u}\|^2)$$

Substituting

$$\|\mathbf{u}\|^2 = u_1^2 + u_2^2 + u_3^2, \qquad \|\mathbf{v}\|^2 = v_1^2 + v_2^2 + v_3^2$$

and

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

we obtain, after simplifying,

$$\boxed{\mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3} \tag{15}$$

The companion formula for vectors in 2-space is

$$\boxed{\mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2} \tag{16}$$

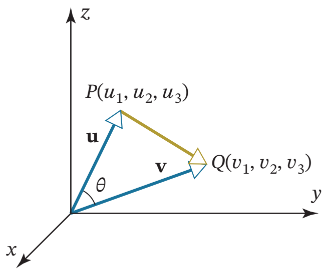


FIGURE 3.2.6

Remark Although we derived Formula (15) and its 2-space companion under the assumption that $\mathbf{u}$ and $\mathbf{v}$ are nonzero, it turned out that these formulas are also applicable if $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$ (verify).

Motivated by the pattern in Formulas (15) and (16), we make the following definition.

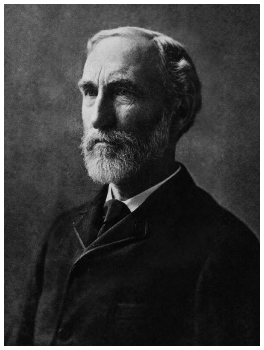
Definition 4

If $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ are vectors in R^n, then the ***dot product*** (also called the ***Euclidean inner product***) of $\mathbf{u}$ and $\mathbf{v}$ is denoted by $\mathbf{u}\cdot\mathbf{v}$ and is defined by

$$\boxed{\mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n} \tag{17}$$

In words, to calculate a dot product multiply corresponding components and add the resulting products.

Historical Note



Josiah Willard Gibbs (1839–1903)

The dot product notation was first introduced by the American physicist and mathematician J. Willard Gibbs in a pamphlet distributed to his students at Yale University in the 1880s. The product was originally written on the baseline, rather than centered as today, and was referred to as the *direct product*. Gibbs's pamphlet was eventually incorporated into a book entitled *Vector Analysis* that was published in 1901 and coauthored with one of his students. Gibbs made major contributions to the fields of thermodynamics and electromagnetic theory and is generally regarded as the greatest American physicist of the nineteenth century.

[Image: Wikipedia Commons]

EXAMPLE 6 | Calculating Dot Products Using Components

(*a*) Use Formula (15) to compute the dot product of the vectors **u** and **v** in Example 5.

(*b*) Calculate $\mathbf{u} \cdot \mathbf{v}$ for the following vectors in R^4:

$$\mathbf{u} = (-1, 3, 5, 7), \quad \mathbf{v} = (-3, -4, 1, 0)$$

Solution (*a*) The component forms of the vectors are $\mathbf{u} = (0, 0, 1)$ and $\mathbf{v} = (0, 2, 2)$. Thus,

$$\mathbf{u} \cdot \mathbf{v} = (0)(0) + (0)(2) + (1)(2) = 2$$

which agrees with the result obtained geometrically in Example 5.

Solution (*b*)

$$\mathbf{u} \cdot \mathbf{v} = (-1)(-3) + (3)(-4) + (5)(1) + (7)(0) = -4$$

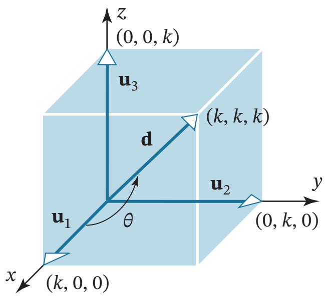


FIGURE 3.2.7

Note that the angle θ obtained in Example 7 does not involve k. Why was this to be expected?

EXAMPLE 7 | A Geometry Problem Solved Using Dot Product

Find the angle between a diagonal of a cube and one of its edges.

Solution Let k be the length of an edge and introduce a coordinate system as shown in Figure 3.2.7. If we let $\mathbf{u}_1 = (k, 0, 0)$, $\mathbf{u}_2 = (0, k, 0)$, and $\mathbf{u}_3 = (0, 0, k)$, then the vector

$$\mathbf{d} = (k, k, k) = \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3$$

is a diagonal of the cube. It follows from Formula (13) that the angle θ between **d** and the edge $\mathbf{u}_1$ satisfies

$$\cos\theta = \frac{\mathbf{u}_1 \cdot \mathbf{d}}{\|\mathbf{u}_1\|\|\mathbf{d}\|} = \frac{k^2}{(k)(\sqrt{3k^2})} = \frac{1}{\sqrt{3}}$$

With the help of a calculator we obtain

$$\theta = \cos^{-1}\left(\frac{1}{\sqrt{3}}\right) \approx 54.74°$$

Algebraic Properties of the Dot Product

In the special case where $\mathbf{u} = \mathbf{v}$ in Definition 4, we obtain the relationship

$$\mathbf{v} \cdot \mathbf{v} = v_1^2 + v_2^2 + \cdots + v_n^2 = \|\mathbf{v}\|^2 \tag{18}$$

This yields the following formula for expressing the length of a vector in terms of a dot product:

$$\boxed{\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}} \tag{19}$$

Dot products have many of the same algebraic properties as products of real numbers.

Theorem 3.2.2

If **u**, **v**, and **w** are vectors in R^n, and if k is a scalar, then:

(*a*) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$ **[Symmetry property]**

(*b*) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$ **[Distributive property]**

(*c*) $k(\mathbf{u} \cdot \mathbf{v}) = (k\mathbf{u}) \cdot \mathbf{v}$ **[Homogeneity property]**

(*d*) $\mathbf{v} \cdot \mathbf{v} \geq 0$ and $\mathbf{v} \cdot \mathbf{v} = 0$ if and only if $\mathbf{v} = \mathbf{0}$ **[Positivity property]**

We will prove parts (*c*) and (*d*) and leave the other proofs as exercises.

Proof (c) Let $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$. Then

$$\begin{aligned} k(\mathbf{u}\cdot\mathbf{v}) &= k(u_1v_1 + u_2v_2 + \cdots + u_nv_n) \\ &= (ku_1)v_1 + (ku_2)v_2 + \cdots + (ku_n)v_n = (k\mathbf{u})\cdot\mathbf{v} \end{aligned}$$

Proof (d) The result follows from parts (*a*) and (*b*) of Theorem 3.2.1 and the fact that

$$\mathbf{v}\cdot\mathbf{v} = v_1v_1 + v_2v_2 + \cdots + v_nv_n = v_1^2 + v_2^2 + \cdots + v_n^2 = \|\mathbf{v}\|^2 \quad ■$$

The next theorem gives additional properties of dot products. The proofs can be obtained either by expressing the vectors in terms of components or by using the algebraic properties established in Theorem 3.2.2.

Theorem 3.2.3

If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in R^n, and if k is a scalar, then:

(*a*) $\mathbf{0}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{0} = 0$

(*b*) $(\mathbf{u}+\mathbf{v})\cdot\mathbf{w} = \mathbf{u}\cdot\mathbf{w} + \mathbf{v}\cdot\mathbf{w}$

(*c*) $\mathbf{u}\cdot(\mathbf{v}-\mathbf{w}) = \mathbf{u}\cdot\mathbf{v} - \mathbf{u}\cdot\mathbf{w}$

(*d*) $(\mathbf{u}-\mathbf{v})\cdot\mathbf{w} = \mathbf{u}\cdot\mathbf{w} - \mathbf{v}\cdot\mathbf{w}$

(*e*) $k(\mathbf{u}\cdot\mathbf{v}) = \mathbf{u}\cdot(k\mathbf{v})$

We will show how Theorem 3.2.2 can be used to prove part (*b*) without breaking the vectors into components. The other proofs are left as exercises.

Proof (b)

$$\begin{aligned} (\mathbf{u}+\mathbf{v})\cdot\mathbf{w} &= \mathbf{w}\cdot(\mathbf{u}+\mathbf{v}) && \textbf{[By symmetry]} \\ &= \mathbf{w}\cdot\mathbf{u} + \mathbf{w}\cdot\mathbf{v} && \textbf{[By distributivity]} \\ &= \mathbf{u}\cdot\mathbf{w} + \mathbf{v}\cdot\mathbf{w} && \textbf{[By symmetry]} \quad ■ \end{aligned}$$

Formulas (18) and (19) together with Theorems 3.2.2 and 3.2.3 make it possible to manipulate expressions involving dot products using familiar algebraic techniques.

EXAMPLE 8 | Calculating with Dot Products

$$\begin{aligned} (\mathbf{u}-2\mathbf{v})\cdot(3\mathbf{u}+4\mathbf{v}) &= \mathbf{u}\cdot(3\mathbf{u}+4\mathbf{v}) - 2\mathbf{v}\cdot(3\mathbf{u}+4\mathbf{v}) \\ &= 3(\mathbf{u}\cdot\mathbf{u}) + 4(\mathbf{u}\cdot\mathbf{v}) - 6(\mathbf{v}\cdot\mathbf{u}) - 8(\mathbf{v}\cdot\mathbf{v}) \\ &= 3\|\mathbf{u}\|^2 - 2(\mathbf{u}\cdot\mathbf{v}) - 8\|\mathbf{v}\|^2 \end{aligned}$$

Cauchy–Schwarz Inequality and Angles in R^n

Our next objective is to extend to R^n the notion of "angle" between nonzero vectors $\mathbf{u}$ and $\mathbf{v}$. We will do this by starting with the formula

$$\boxed{\theta = \cos^{-1}\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}\right)} \tag{20}$$

which follows from Formula (13) that we previously derived for nonzero vectors in R^2 and R^3. Since dot products and norms have been defined for vectors in R^n, it would seem that this formula has all the ingredients to serve as a *definition* of the angle θ between two vectors, $\mathbf{u}$ and $\mathbf{v}$, in R^n. However, there is a fly in the ointment, the problem being that this formula is not valid unless its argument satisfies the inequalities

$$-1 \leq \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|} \leq 1 \tag{21}$$

Fortunately, these inequalities *do* hold for all nonzero vectors in R^n as a result of the following fundamental result known as the ***Cauchy–Schwarz inequality***.

Theorem 3.2.4

Cauchy–Schwarz Inequality

If $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$ are vectors in R^n, then

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\|\,\|\mathbf{v}\| \tag{22}$$

or in terms of components

$$|u_1v_1 + u_2v_2 + \cdots + u_nv_n| \leq (u_1^2 + u_2^2 + \cdots + u_n^2)^{1/2}(v_1^2 + v_2^2 + \cdots + v_n^2)^{1/2} \tag{23}$$

We will omit the proof of this theorem because later in the text we will prove a more general version of which this will be a special case. Our goal for now will be to use this theorem to prove that the inequalities in (21) hold for all nonzero vectors in R^n. Once that is done we will have established all the results required to use Formula (20) as our *definition* of the angle between nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in R^n.

Historical Note

Hermann Amandus Schwarz (1843–1921)

Viktor Yakovlevich Bunyakovsky (1804–1889)

The Cauchy–Schwarz inequality is named in honor of the French mathematician Augustin Cauchy (see p. 136) and the German mathematician Hermann Schwarz. Variations of this inequality occur in many different settings and under various names. Depending on the context in which the inequality occurs, you may find it called Cauchy's inequality, the Schwarz inequality, or sometimes even the Bunyakovsky inequality, in recognition of the Russian mathematician who published his version of the inequality in 1859, about 25 years before Schwarz.

[Images: Ludwig Zipfel/Wikipedia Common (Schwarz); University of St-Andrews/Wikipedia (Bunyakovsky)]

To prove that the inequalities in (21) hold for all nonzero vectors in R^n, divide both sides of Formula (22) by the product $\|\mathbf{u}\|\,\|\mathbf{v}\|$ to obtain

$$\frac{|\mathbf{u}\cdot\mathbf{v}|}{\|\mathbf{u}\|\,\|\mathbf{v}\|} \le 1 \quad \text{or equivalently} \quad \left|\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}\right| \le 1$$

from which (21) follows.

Geometry in R^n

Our next theorem will extend two familiar plane geometry results to R^n: the sum of the lengths of two sides of a triangle is at least as large as the third side (**Figure 3.2.8**), and the shortest distance between two points is a straight line (**Figure 3.2.9**).

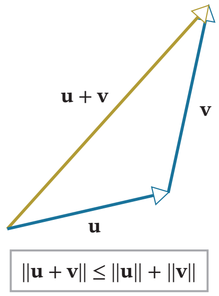


FIGURE 3.2.8

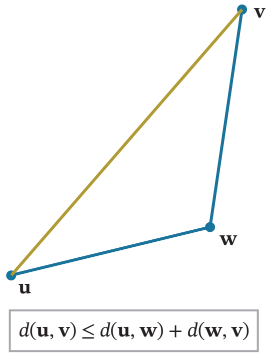


FIGURE 3.2.9

Theorem 3.2.5

If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in R^n, then:

(*a*) $\|\mathbf{u}+\mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|$ **[Triangle inequality for vectors]**

(*b*) $d(\mathbf{u},\mathbf{v}) \le d(\mathbf{u},\mathbf{w}) + d(\mathbf{w},\mathbf{v})$ **[Triangle inequality for distances]**

***Proof* (*a*)**

$$\begin{aligned}
\|\mathbf{u}+\mathbf{v}\|^2 &= (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) = (\mathbf{u}\cdot\mathbf{u}) + 2(\mathbf{u}\cdot\mathbf{v}) + (\mathbf{v}\cdot\mathbf{v}) \\
&= \|\mathbf{u}\|^2 + 2(\mathbf{u}\cdot\mathbf{v}) + \|\mathbf{v}\|^2 \\
&\le \|\mathbf{u}\|^2 + 2|\mathbf{u}\cdot\mathbf{v}| + \|\mathbf{v}\|^2 && \longleftarrow \text{Property of absolute value} \\
&\le \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\,\|\mathbf{v}\| + \|\mathbf{v}\|^2 && \longleftarrow \text{Cauchy–Schwarz inequality} \\
&= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2 && \longleftarrow \text{Algebraic simplification}
\end{aligned}$$

This completes the proof since both sides of the inequality in part (*a*) are nonnegative.

***Proof* (*b*)** It follows from part (*a*) and Formula (11) that

$$\begin{aligned}
d(\mathbf{u},\mathbf{v}) &= \|\mathbf{u}-\mathbf{v}\| = \|(\mathbf{u}-\mathbf{w}) + (\mathbf{w}-\mathbf{v})\| \\
&\le \|\mathbf{u}-\mathbf{w}\| + \|\mathbf{w}-\mathbf{v}\| = d(\mathbf{u},\mathbf{w}) + d(\mathbf{w},\mathbf{v}) \quad \blacksquare
\end{aligned}$$

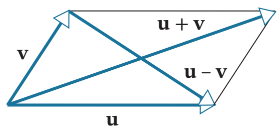


FIGURE 3.2.10

It is proved in plane geometry that for any parallelogram the sum of the squares of the diagonals is equal to the sum of the squares of the four sides (**Figure 3.2.10**). The following theorem generalizes that result to R^n.

Theorem 3.2.6

Parallelogram Equation for Vectors

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in R^n, then

$$\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\left(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2\right) \tag{24}$$

Proof

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

We could state and prove many more theorems from plane geometry that generalize to R^n, but the ones already given should suffice to convince you that R^n is not so different from R^2 and R^3 even though we cannot visualize it directly. The next theorem establishes a fundamental relationship between the dot product and norm in R^n.

Theorem 3.2.7

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in R^n with the Euclidean inner product, then

$$\mathbf{u}\cdot\mathbf{v} = \tfrac{1}{4}\|\mathbf{u}+\mathbf{v}\|^2 - \tfrac{1}{4}\|\mathbf{u}-\mathbf{v}\|^2 \tag{25}$$

Proof

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

from which (25) follows by simple algebra. $\blacksquare$

Application of Dot Products to ISBN Numbers

Although the system changed in 2007, most older books have been assigned a unique 10-digit number called an ***International Standard Book Number*** or ISBN. The first nine digits of this number are split into three groups—the first group representing the country or group of countries in which the book originates, the second identifying the publisher, and the third assigned to the book title itself. The tenth and final digit, called a ***check digit***, is computed from the first nine digits and is used to ensure that an electronic transmission of the ISBN, say over the Internet, occurs without error.

To explain how this is done, regard the first nine digits of the ISBN as a vector $\mathbf{b}$ in R^9, and let $\mathbf{a}$ be the vector

$$\mathbf{a} = (1, 2, 3, 4, 5, 6, 7, 8, 9)$$

Then the check digit c is computed using the following procedure:

1. Form the dot product $\mathbf{a}\cdot\mathbf{b}$.
2. Divide $\mathbf{a}\cdot\mathbf{b}$ by 11, thereby producing a remainder c that is an integer between 0 and 10, inclusive. The check digit is taken to be c, with the proviso that $c = 10$ is written as X to avoid double digits.

For example, the ISBN of the brief edition of *Calculus*, sixth edition, by Howard Anton is

0-471-15307-9

which has a check digit of 9. This is consistent with the first nine digits of the ISBN, since

$$\mathbf{a}\cdot\mathbf{b} = (1, 2, 3, 4, 5, 6, 7, 8, 9)\cdot(0, 4, 7, 1, 1, 5, 3, 0, 7) = 152$$

Dividing 152 by 11 produces a quotient of 13 and a remainder of 9, so the check digit is $c = 9$. If an electronic order is placed for a book with a certain ISBN, then the warehouse can use the above procedure to verify that the check digit is consistent with the first nine digits, thereby reducing the possibility of a costly shipping error.

Dot Products as Matrix Multiplication

There are various ways to express the dot product of vectors using matrix notation. The formulas depend on whether the vectors are expressed as row matrices or column matrices. **Table 1** shows the possibilities.

If A is an $n \times n$ matrix and $\mathbf{u}$ and $\mathbf{v}$ are $n \times 1$ matrices, then it follows from the first row in Table 1 and properties of the transpose that

$$A\mathbf{u}\cdot\mathbf{v} = \mathbf{v}^T(A\mathbf{u}) = (\mathbf{v}^TA)\mathbf{u} = (A^T\mathbf{v})^T\mathbf{u} = \mathbf{u}\cdot A^T\mathbf{v}$$
$$\mathbf{u}\cdot A\mathbf{v} = (A\mathbf{v})^T\mathbf{u} = (\mathbf{v}^TA^T)\mathbf{u} = \mathbf{v}^T(A^T\mathbf{u}) = A^T\mathbf{u}\cdot\mathbf{v}$$

TABLE 1

Form	Dot Product	Example	
$\mathbf{u}$ a column matrix and $\mathbf{v}$ a column matrix	$\mathbf{u}\cdot\mathbf{v}=\mathbf{u}^T\mathbf{v}=\mathbf{v}^T\mathbf{u}$	$\mathbf{u}=\begin{bmatrix}1\\-3\\5\end{bmatrix}$ $\mathbf{v}=\begin{bmatrix}5\\4\\0\end{bmatrix}$	$\mathbf{u}^T\mathbf{v}=[1\quad -3\quad 5]\begin{bmatrix}5\\4\\0\end{bmatrix}=-7$ $\mathbf{v}^T\mathbf{u}=[5\quad 4\quad 0]\begin{bmatrix}1\\-3\\5\end{bmatrix}=-7$
$\mathbf{u}$ a row matrix and $\mathbf{v}$ a column matrix	$\mathbf{u}\cdot\mathbf{v}=\mathbf{u}\mathbf{v}=\mathbf{v}^T\mathbf{u}^T$	$\mathbf{u}=[1\quad -3\quad 5]$ $\mathbf{v}=\begin{bmatrix}5\\4\\0\end{bmatrix}$	$\mathbf{u}\mathbf{v}=[1\quad -3\quad 5]\begin{bmatrix}5\\4\\0\end{bmatrix}=-7$ $\mathbf{v}^T\mathbf{u}^T=[5\quad 4\quad 0]\begin{bmatrix}1\\-3\\5\end{bmatrix}=-7$
$\mathbf{u}$ a column matrix and $\mathbf{v}$ a row matrix	$\mathbf{u}\cdot\mathbf{v}=\mathbf{v}\mathbf{u}=\mathbf{u}^T\mathbf{v}^T$	$\mathbf{u}=\begin{bmatrix}1\\-3\\5\end{bmatrix}$ $\mathbf{v}=[5\quad 4\quad 0]$	$\mathbf{v}\mathbf{u}=[5\quad 4\quad 0]\begin{bmatrix}1\\-3\\5\end{bmatrix}=-7$ $\mathbf{u}^T\mathbf{v}^T=[1\quad -3\quad 5]\begin{bmatrix}5\\4\\0\end{bmatrix}=-7$
$\mathbf{u}$ a row matrix and $\mathbf{v}$ a row matrix	$\mathbf{u}\cdot\mathbf{v}=\mathbf{u}\mathbf{v}^T=\mathbf{v}\mathbf{u}^T$	$\mathbf{u}=[1\quad -3\quad 5]$ $\mathbf{v}=[5\quad 4\quad 0]$	$\mathbf{u}\mathbf{v}^T=[1\quad -3\quad 5]\begin{bmatrix}5\\4\\0\end{bmatrix}=-7$ $\mathbf{v}\mathbf{u}^T=[5\quad 4\quad 0]\begin{bmatrix}1\\-3\\5\end{bmatrix}=-7$

The resulting formulas

$$\boxed{A\mathbf{u}\cdot\mathbf{v}=\mathbf{u}\cdot A^T\mathbf{v}} \tag{26}$$

$$\boxed{\mathbf{u}\cdot A\mathbf{v}=A^T\mathbf{u}\cdot\mathbf{v}} \tag{27}$$

provide an important link between multiplication by an $n\times n$ matrix A and multiplication by A^T.

EXAMPLE 9 | Verifying that $A\mathbf{u}\cdot\mathbf{v}=\mathbf{u}\cdot A^T\mathbf{v}$

Suppose that

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

Then

$$A\mathbf{u}=\begin{bmatrix}1&-2&3\\2&4&1\\-1&0&1\end{bmatrix}\begin{bmatrix}-1\\2\\4\end{bmatrix}=\begin{bmatrix}7\\10\\5\end{bmatrix}$$

$$A^T\mathbf{v}=\begin{bmatrix}1&2&-1\\-2&4&0\\3&1&1\end{bmatrix}\begin{bmatrix}-2\\0\\5\end{bmatrix}=\begin{bmatrix}-7\\4\\-1\end{bmatrix}$$

from which we obtain

$$A\mathbf{u}\cdot\mathbf{v} = 7(-2) + 10(0) + 5(5) = 11$$
$$\mathbf{u}\cdot A^T\mathbf{v} = (-1)(-7) + 2(4) + 4(-1) = 11$$

Thus, $A\mathbf{u}\cdot\mathbf{v} = \mathbf{u}\cdot A^T\mathbf{v}$ as guaranteed by Formula (26). We leave it for you to verify that Formula (27) also holds.

A Dot Product View of Matrix Multiplication

Dot products provide another way of thinking about matrix multiplication. Recall that if $A = [a_{ij}]$ is an $m \times r$ matrix and $B = [b_{ij}]$ is an $r \times n$ matrix, then by the row-column rule stated in Formula (5) of Section 1.3 the ijth entry of AB is

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ir}b_{rj}$$

which is the dot product of the ith row vector of A

$$[a_{i1} \quad a_{i2} \quad \cdots \quad a_{ir}]$$

and the jth column vector of B

$$\begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{rj} \end{bmatrix}$$

Thus, if we denote the row vectors of A by $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m$ and the column vectors of the matrix B by $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$, then the matrix product AB can be expressed as

$$AB = \begin{bmatrix} \mathbf{r}_1\cdot\mathbf{c}_1 & \mathbf{r}_1\cdot\mathbf{c}_2 & \cdots & \mathbf{r}_1\cdot\mathbf{c}_n \\ \mathbf{r}_2\cdot\mathbf{c}_1 & \mathbf{r}_2\cdot\mathbf{c}_2 & \cdots & \mathbf{r}_2\cdot\mathbf{c}_n \\ \vdots & \vdots & & \vdots \\ \mathbf{r}_m\cdot\mathbf{c}_1 & \mathbf{r}_m\cdot\mathbf{c}_2 & \cdots & \mathbf{r}_m\cdot\mathbf{c}_n \end{bmatrix} \tag{28}$$

Exercise Set 3.2

In Exercises 1–2, find the norm of $\mathbf{v}$*, and a unit vector that is oppositely directed to* $\mathbf{v}$.

1. **a.** $\mathbf{v} = (2, 2, 2)$ **b.** $\mathbf{v} = (1, 0, 2, 1, 3)$

2. **a.** $\mathbf{v} = (1, -1, 2)$ **b.** $\mathbf{v} = (-2, 3, 3, -1)$

In Exercises 3–4, evaluate the given expression with $\mathbf{u} = (2, -2, 3)$, $\mathbf{v} = (1, -3, 4)$, *and* $\mathbf{w} = (3, 6, -4)$.

3. **a.** $\|\mathbf{u} + \mathbf{v}\|$ **b.** $\|\mathbf{u}\| + \|\mathbf{v}\|$

c. $\|-2\mathbf{u} + 2\mathbf{v}\|$ **d.** $\|3\mathbf{u} - 5\mathbf{v} + \mathbf{w}\|$

4. **a.** $\|\mathbf{u} + \mathbf{v} + \mathbf{w}\|$ **b.** $\|\mathbf{u} - \mathbf{v}\|$

c. $\|3\mathbf{v}\| - 3\|\mathbf{v}\|$ **d.** $\|\mathbf{u}\| - \|\mathbf{v}\|$

In Exercises 5–6, evaluate the given expression with $\mathbf{u} = (-2, -1, 4, 5)$, $\mathbf{v} = (3, 1, -5, 7)$, *and* $\mathbf{w} = (-6, 2, 1, 1)$.

5. **a.** $\|3\mathbf{u} - 5\mathbf{v} + \mathbf{w}\|$ **b.** $\|3\mathbf{u}\| - 5\|\mathbf{v}\| + \|\mathbf{w}\|$

c. $\big\|-\|\mathbf{u}\|\mathbf{v}\big\|$

6. **a.** $\|\mathbf{u}\| + \|-2\mathbf{v}\| + \|-3\mathbf{w}\|$ **b.** $\big\|\|\mathbf{u} - \mathbf{v}\|\mathbf{w}\big\|$

7. Let $\mathbf{v} = (-2, 3, 0, 6)$. Find all scalars k such that $\|k\mathbf{v}\| = 5$.

8. Let $\mathbf{v} = (1, 1, 2, -3, 1)$. Find all scalars k such that $\|k\mathbf{v}\| = 4$.

In Exercises 9–10, find $\mathbf{u}\cdot\mathbf{v}$, $\mathbf{u}\cdot\mathbf{u}$, *and* $\mathbf{v}\cdot\mathbf{v}$.

9. **a.** $\mathbf{u} = (3, 1, 4)$, $\mathbf{v} = (2, 2, -4)$

b. $\mathbf{u} = (1, 1, 4, 6)$, $\mathbf{v} = (2, -2, 3, -2)$

10. **a.** $\mathbf{u} = (1, 1, -2, 3)$, $\mathbf{v} = (-1, 0, 5, 1)$

b. $\mathbf{u} = (2, -1, 1, 0, -2)$, $\mathbf{v} = (1, 2, 2, 2, 1)$

In Exercises 11–12, find the Euclidean distance between $\mathbf{u}$ *and* $\mathbf{v}$ *and the cosine of the angle between those vectors. State whether that angle is acute, obtuse, or 90°.*

11. **a.** $\mathbf{u} = (3, 3, 3)$, $\mathbf{v} = (1, 0, 4)$

b. $\mathbf{u} = (0, -2, -1, 1)$, $\mathbf{v} = (-3, 2, 4, 4)$

12. **a.** $\mathbf{u} = (1, 2, -3, 0)$, $\mathbf{v} = (5, 1, 2, -2)$

b. $\mathbf{u} = (0, 1, 1, 1, 2)$, $\mathbf{v} = (2, 1, 0, -1, 3)$

13. Suppose that a vector $\mathbf{a}$ in the xy-plane has a length of 9 units and points in a direction that is 120° counterclockwise from the positive x-axis, and a vector $\mathbf{b}$ in that plane has a length of 5 units and points in the positive y-direction. Find $\mathbf{a}\cdot\mathbf{b}$.

14. Suppose that a vector **a** in the *xy*-plane points in a direction that is 47° counterclockwise from the positive *x*-axis, and a vector **b** in that plane points in a direction that is 43° clockwise from the positive *x*-axis. What can you say about the value of $\mathbf{a} \cdot \mathbf{b}$?

In Exercises 15–16, determine whether the expression makes sense mathematically. If not, explain why.

15. a. $\mathbf{u} \cdot (\mathbf{v} \cdot \mathbf{w})$ **b.** $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w})$

c. $\|\mathbf{u} \cdot \mathbf{v}\|$ **d.** $(\mathbf{u} \cdot \mathbf{v}) - \|\mathbf{u}\|$

16. a. $\|\mathbf{u}\| \cdot \|\mathbf{v}\|$ **b.** $(\mathbf{u} \cdot \mathbf{v}) - \mathbf{w}$

c. $(\mathbf{u} \cdot \mathbf{v}) - k$ **d.** $k \cdot \mathbf{u}$

In Exercises 17–18, verify that the Cauchy–Schwarz inequality holds.

17. a. $\mathbf{u} = (-3, 1, 0),\ \mathbf{v} = (2, -1, 3)$

b. $\mathbf{u} = (0, 2, 2, 1),\ \mathbf{v} = (1, 1, 1, 1)$

18. a. $\mathbf{u} = (4, 1, 1),\ \mathbf{v} = (1, 2, 3)$

b. $\mathbf{u} = (1, 2, 1, 2, 3),\ \mathbf{v} = (0, 1, 1, 5, -2)$

19. Let $\mathbf{r}_0 = (x_0, y_0)$ be a fixed vector in R^2. In each part, describe in words the set of all vectors $\mathbf{r} = (x, y)$ that satisfy the stated condition.

a. $\|\mathbf{r} - \mathbf{r}_0\| = 1$ **b.** $\|\mathbf{r} - \mathbf{r}_0\| \le 1$ **c.** $\|\mathbf{r} - \mathbf{r}_0\| > 1$

20. Repeat the directions of Exercise 19 for vectors $\mathbf{r} = (x, y, z)$ and $\mathbf{r}_0 = (x_0, y_0, z_0)$ in R^3.

Exercises 21–25 *The direction of a nonzero vector* **v** *in an xyz-coordinate system is completely determined by the angles α, β, and γ between* **v** *and the standard unit vectors* **i**, **j**, *and* **k** *(Figure Ex-21). These are called the* **direction angles** *of* **v**, *and their cosines are called the* **direction cosines** *of* **v**.

21. Use Formula (13) to show that the direction cosines of a vector $\mathbf{v} = (v_1, v_2, v_3)$ in R^3 are

$$\cos\alpha = \frac{v_1}{\|\mathbf{v}\|}, \quad \cos\beta = \frac{v_2}{\|\mathbf{v}\|}, \quad \cos\gamma = \frac{v_3}{\|\mathbf{v}\|}$$

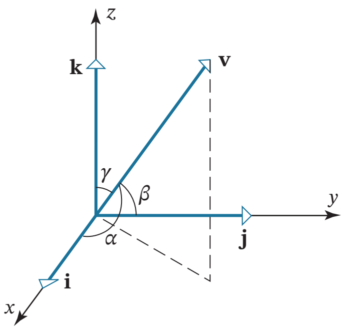


FIGURE Ex-21

22. Use the result in Exercise 21 to show that

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$$

23. Show that two nonzero vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in R^3 are orthogonal if and only if their direction cosines satisfy

$$\cos\alpha_1 \cos\alpha_2 + \cos\beta_1 \cos\beta_2 + \cos\gamma_1 \cos\gamma_2 = 0$$

24. The accompanying figure shows a cube.

a. Find the angle between the vectors **d** and **u** to the nearest degree.

b. Make a conjecture about the angle between the vectors **d** and **v**, and confirm your conjecture by computing the angle.

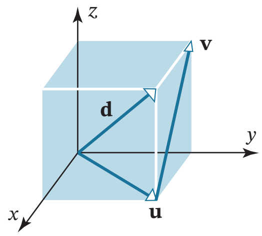


FIGURE Ex-24

25. Estimate, to the nearest degree, the angles that a diagonal of a box with dimensions 10 cm × 15 cm × 25 cm makes with the edges of the box.

26. If $\|\mathbf{v}\| = 2$ and $\|\mathbf{w}\| = 3$, what are the largest and smallest values possible for $\|\mathbf{v} - \mathbf{w}\|$? Give a geometric explanation of your results.

27. What can you say about two nonzero vectors, **u** and **v**, that satisfy the equation $\|\mathbf{u} + \mathbf{v}\| = \|\mathbf{u}\| + \|\mathbf{v}\|$?

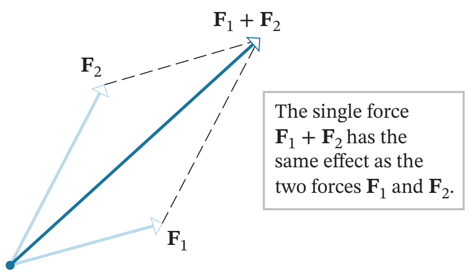
28. a. What relationship must hold for the point $\mathbf{p} = (a, b, c)$ to be equidistant from the origin and the *xz*-plane? Make sure that the relationship you state is valid for positive and negative values of *a*, *b*, and *c*.

b. What relationship must hold for the point $\mathbf{p} = (a, b, c)$ to be farther from the origin than from the *xz*-plane? Make sure that the relationship you state is valid for positive and negative values of *a*, *b*, and *c*.

29. State a procedure for finding a vector of a specified length *m* that points in the same direction as a given vector **v**.

30. Under what conditions will the triangle inequality (Theorem 3.2.5*a*) be an equality? Explain your answer geometrically.

Exercises 31–32 *The effect that a force has on an object depends on the magnitude of the force and the direction in which it is applied. Thus, forces can be regarded as vectors and represented as arrows in which the length of the arrow specifies the magnitude of the force, and the direction of the arrow specifies the direction in which the force is applied. It is a fact of physics that force vectors obey the parallelogram law in the sense that if two force vectors* $\mathbf{F}_1$ *and* $\mathbf{F}_2$ *are applied at a point on an object, then the effect is the same as if the single force* $\mathbf{F}_1 + \mathbf{F}_2$ *(called the* **resultant***) were applied at that point (see accompanying figure). Forces are commonly measured in units called pounds-force (abbreviated lbf) or Newtons (abbreviated N).*



31. A particle is said to be in ***static equilibrium*** if the resultant of all forces applied to it is zero. For the forces in the accompanying figure, find the resultant **F** that must be applied to the indicated point to produce static equilibrium. Describe **F** by giving its magnitude and the angle in degrees that it makes with the positive x-axis.

32. Follow the directions of Exercise 31.

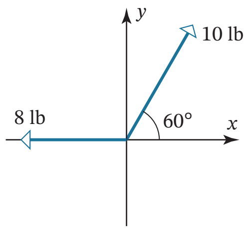


FIGURE Ex-31

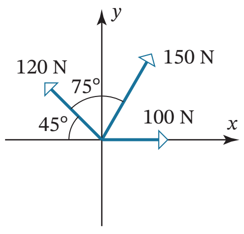


FIGURE Ex-32

Working with Proofs

33. Prove parts (a) and (b) of Theorem 3.2.1.

34. Prove parts (a) and (c) of Theorem 3.2.3.

35. Prove parts (d) and (e) of Theorem 3.2.3.

True-False Exercises

TF. In parts **(a)–(j)** determine whether the statement is true or false, and justify your answer.

a. If each component of a vector in R^3 is doubled, the norm of that vector is doubled.

b. In R^2, the vectors of norm 5 whose initial points are at the origin have terminal points lying on a circle of radius 5 centered at the origin.

c. Every vector in R^n has a positive norm.

d. If $\mathbf{v}$ is a nonzero vector in R^n, there are exactly two unit vectors that are parallel to $\mathbf{v}$.

e. If $\|\mathbf{u}\| = 2$, $\|\mathbf{v}\| = 1$, and $\mathbf{u}\cdot\mathbf{v} = 1$, then the angle between $\mathbf{u}$ and $\mathbf{v}$ is $\pi/3$ radians.

f. The expressions $(\mathbf{u}\cdot\mathbf{v}) + \mathbf{w}$ and $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})$ are both meaningful and equal to each other.

g. If $\mathbf{u}\cdot\mathbf{v} = \mathbf{u}\cdot\mathbf{w}$, then $\mathbf{v} = \mathbf{w}$.

h. If $\mathbf{u}\cdot\mathbf{v} = 0$, then either $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$.

i. In R^2, if $\mathbf{u}$ lies in the first quadrant and $\mathbf{v}$ lies in the third quadrant, then $\mathbf{u}\cdot\mathbf{v}$ cannot be positive.

j. For all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in R^n, we have

$$\|\mathbf{u}+\mathbf{v}+\mathbf{w}\| \le \|\mathbf{u}\| + \|\mathbf{v}\| + \|\mathbf{w}\|$$

Working with Technology

T1. Let $\mathbf{u}$ be a vector in R^{100} whose ith component is i, and let $\mathbf{v}$ be the vector in R^{100} whose ith component is $1/(i+1)$. Find the dot product of $\mathbf{u}$ and $\mathbf{v}$.

T2. Find, to the nearest degree, the angles that a diagonal of a box with dimensions 10 cm × 11 cm × 25 cm makes with the edges of the box.

3.3 Orthogonality

In the last section we defined the notion of "angle" between vectors in R^n. In this section we will focus on the notion of "perpendicularity." Perpendicular vectors in R^n play an important role in a wide variety of applications.

Orthogonal Vectors

Recall from Formula (20) in the previous section that the angle θ between two *nonzero* vectors $\mathbf{u}$ and $\mathbf{v}$ in R^n is defined by the formula

$$\theta = \cos^{-1}\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}\right)$$

It follows from this that $\theta = \pi/2$ if and only if $\mathbf{u}\cdot\mathbf{v} = 0$. Thus, we make the following definition.

Definition 1

Two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in R^n are said to be ***orthogonal*** (or ***perpendicular***) if $\mathbf{u} \cdot \mathbf{v} = 0$. We will also agree that the zero vector in R^n is orthogonal to *every* vector in R^n.

EXAMPLE 1 | Orthogonal Vectors

(*a*) Show that $\mathbf{u} = (-2, 3, 1, 4)$ and $\mathbf{v} = (1, 2, 0, -1)$ are orthogonal vectors in R^4.

(*b*) Let $S = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ be the set of standard unit vectors in R^3. Show that each ordered pair of vectors in S is orthogonal.

Solution (*a*) The vectors are orthogonal since

$$\mathbf{u} \cdot \mathbf{v} = (-2)(1) + (3)(2) + (1)(0) + (4)(-1) = 0$$

Solution (*b*) It suffices to show that

$$\mathbf{i} \cdot \mathbf{j} = \mathbf{i} \cdot \mathbf{k} = \mathbf{j} \cdot \mathbf{k} = \mathbf{0}$$

because it will follow automatically from the symmetry property of the dot product that

$$\mathbf{j} \cdot \mathbf{i} = \mathbf{k} \cdot \mathbf{i} = \mathbf{k} \cdot \mathbf{j} = \mathbf{0}$$

Although the orthogonality of the vectors in S is evident geometrically from Figure 3.2.2, it is confirmed algebraically by the computations

$$\mathbf{i} \cdot \mathbf{j} = (1, 0, 0) \cdot (0, 1, 0) = 0$$
$$\mathbf{i} \cdot \mathbf{k} = (1, 0, 0) \cdot (0, 0, 1) = 0$$
$$\mathbf{j} \cdot \mathbf{k} = (0, 1, 0) \cdot (0, 0, 1) = 0$$

Using the computations in R^3 as a model, you should be able to see that each ordered pair of standard unit vectors in R^n is orthogonal.

Lines and Planes Determined by Points and Normals

One learns in analytic geometry that a line in R^2 is determined uniquely by its slope and one of its points, and that a plane in R^3 is determined uniquely by its "inclination" and one of its points. One way of specifying slope and inclination is to use a *nonzero* vector $\mathbf{n}$, called a ***normal***, that is orthogonal to the line or plane in question. For example, **Figure 3.3.1** shows the line through the point $P_0(x_0, y_0)$ that has normal $\mathbf{n} = (a, b)$ and the plane through the point $P_0(x_0, y_0, z_0)$ that has normal $\mathbf{n} = (a, b, c)$. Both the line and the plane are represented by the vector equation

$$\mathbf{n} \cdot \overrightarrow{P_0P} = 0 \tag{1}$$

where P is either an arbitrary point (x, y) on the line or an arbitrary point (x, y, z) in the plane. The vector $\overrightarrow{P_0P}$ can be expressed in terms of components as

$$\overrightarrow{P_0P} = (x - x_0, y - y_0) \quad \textbf{[line]}$$
$$\overrightarrow{P_0P} = (x - x_0, y - y_0, z - z_0) \quad \textbf{[plane]}$$

Thus, Equation (1) can be written as

$$\boxed{a(x - x_0) + b(y - y_0) = 0 \quad \textbf{[line]}} \tag{2}$$

$$\boxed{a(x - x_0) + b(y - y_0) + c(z - z_0) = 0 \quad \textbf{[plane]}} \tag{3}$$

These are called the ***point-normal*** equations of the line and plane.

Formula (1) is called the ***point-normal*** form of a line or plane and Formulas (2) and (3) the ***component*** forms.

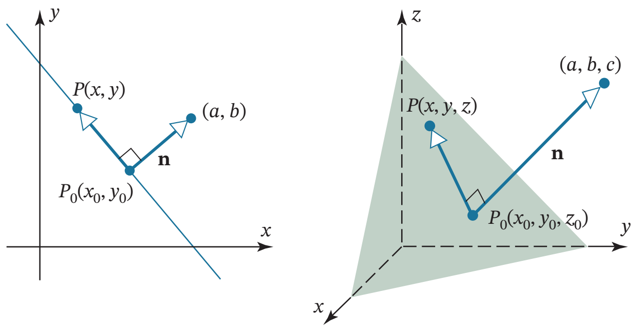


FIGURE 3.3.1

EXAMPLE 2 | Point-Normal Equations

It follows from (2) that in R^2 the equation

$$6(x-3)+(y+7)=0$$

represents the line through the point $(3,-7)$ with normal $\mathbf{n}=(6,1)$; and it follows from (3) that in R^3 the equation

$$4(x-3)+2y-5(z-7)=0$$

represents the plane through the point $(3,0,7)$ with normal $\mathbf{n}=(4,2,-5)$.

When convenient, the terms in Equations (2) and (3) can be multiplied out and the constants combined. This leads to the following theorem.

Theorem 3.3.1

(a) If a and b are constants that are not both zero, then an equation of the form

$$ax+by+c=0 \tag{4}$$

represents a line in R^2 with normal $\mathbf{n}=(a,b)$.

(b) If a, b, and c are constants that are not all zero, then an equation of the form

$$ax+by+cz+d=0 \tag{5}$$

represents a plane in R^3 with normal $\mathbf{n}=(a,b,c)$.

EXAMPLE 3 | Vectors Orthogonal to Lines and Planes Through the Origin

(a) The equation $ax+by=0$ represents a line through the origin in R^2. Show that the vector $\mathbf{n}_1=(a,b)$ formed from the coefficients of the equation is orthogonal to the line, that is, orthogonal to every vector along the line.

(b) The equation $ax+by+cz=0$ represents a plane through the origin in R^3. Show that the vector $\mathbf{n}_2=(a,b,c)$ formed from the coefficients of the equation is orthogonal to the plane, that is, orthogonal to every vector that lies in the plane.

Solution We will solve both problems together. The two equations can be written as

$$(a, b) \cdot (x, y) = 0 \quad \text{and} \quad (a, b, c) \cdot (x, y, z) = 0$$

or, alternatively, as

$$\mathbf{n}_1 \cdot (x, y) = 0 \quad \text{and} \quad \mathbf{n}_2 \cdot (x, y, z) = 0$$

These equations show that $\mathbf{n}_1$ is orthogonal to every vector (x, y) on the line and that $\mathbf{n}_2$ is orthogonal to every vector (x, y, z) in the plane (Figure 3.3.1).

Recall that

$$ax + by = 0 \quad \text{and} \quad ax + by + cz = 0$$

are called *homogeneous equations*. Example 3.3 illustrates that homogeneous equations in two or three unknowns can be written in the vector form

$$\boxed{\mathbf{n} \cdot \mathbf{x} = 0} \tag{6}$$

where $\mathbf{n}$ is the vector of coefficients and $\mathbf{x}$ is the vector of unknowns. In R^2 this is called the ***vector form of a line*** through the origin, and in R^3 it is called the ***vector form of a plane*** through the origin.

Referring to Table 1 of Section 3.2, in what other ways can you write (6) if $\mathbf{n}$ and $\mathbf{x}$ are expressed in matrix form?

Orthogonal Projections

In many applications it is necessary to "decompose" a vector $\mathbf{u}$ into a sum of two terms, one term being a scalar multiple of a specified nonzero vector $\mathbf{a}$ and the other term being orthogonal to $\mathbf{a}$. For example, if $\mathbf{u}$ and $\mathbf{a}$ are vectors in R^2 that are positioned so their initial points coincide at a point Q, then we can create such a decomposition as follows (**Figure 3.3.2**):

- Drop a perpendicular from the tip of $\mathbf{u}$ to the line through $\mathbf{a}$.
- Construct the vector $\mathbf{w}_1$ from Q to the foot of the perpendicular.
- Construct the vector $\mathbf{w}_2 = \mathbf{u} - \mathbf{w}_1$.

Since

$$\mathbf{w}_1 + \mathbf{w}_2 = \mathbf{w}_1 + (\mathbf{u} - \mathbf{w}_1) = \mathbf{u}$$

we have decomposed $\mathbf{u}$ into a sum of two orthogonal vectors, the first term being a scalar multiple of $\mathbf{a}$ and the second being orthogonal to $\mathbf{a}$.

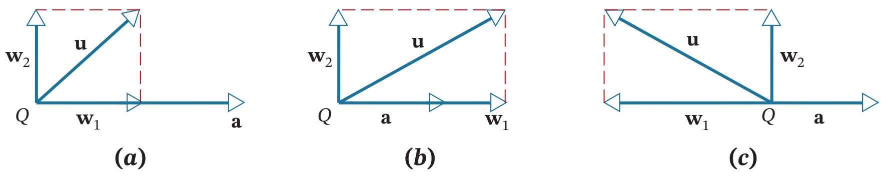


FIGURE 3.3.2 Three possible cases.

The following theorem shows that the foregoing results, which we illustrated using vectors in R^2, apply as well in R^n.

Theorem 3.3.2

Projection Theorem

If $\mathbf{u}$ and $\mathbf{a}$ are vectors in R^n, and if $\mathbf{a} \neq 0$, then $\mathbf{u}$ can be expressed in exactly one way in the form $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1$ is a scalar multiple of $\mathbf{a}$ and $\mathbf{w}_2$ is orthogonal to $\mathbf{a}$.

Proof Since the vector $\mathbf{w}_1$ is to be a scalar multiple of $\mathbf{a}$, it must have the form

$$\mathbf{w}_1 = k\mathbf{a} \tag{7}$$

Our goal is to find a value of the scalar k and a vector $\mathbf{w}_2$ that is orthogonal to $\mathbf{a}$ such that

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2 \tag{8}$$

We can determine k by using (7) to rewrite (8) as

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2 = k\mathbf{a} + \mathbf{w}_2$$

and then applying Theorems 3.2.2 and 3.2.3 to obtain

$$\mathbf{u} \cdot \mathbf{a} = (k\mathbf{a} + \mathbf{w}_2) \cdot \mathbf{a} = k\|\mathbf{a}\|^2 + (\mathbf{w}_2 \cdot \mathbf{a}) \tag{9}$$

Since $\mathbf{w}_2$ is to be orthogonal to $\mathbf{a}$, the last term in (9) must be 0, and hence k must satisfy the equation

$$\mathbf{u} \cdot \mathbf{a} = k\|\mathbf{a}\|^2$$

from which we obtain

$$k = \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}$$

as the only possible value for k. The proof can be completed by rewriting (8) as

$$\mathbf{w}_2 = \mathbf{u} - \mathbf{w}_1 = \mathbf{u} - k\mathbf{a} = \mathbf{u} - \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a}$$

and then confirming that $\mathbf{w}_2$ is orthogonal to $\mathbf{a}$ by showing that $\mathbf{w}_2 \cdot \mathbf{a} = 0$ (we leave the details for you). ■

The vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ in the Projection Theorem have associated names—the vector $\mathbf{w}_1$ is called the ***orthogonal projection of u on a*** or sometimes ***the vector component of u along a***, and the vector $\mathbf{w}_2$ is called the vector ***component of u orthogonal to a***. The vector $\mathbf{w}_1$ is commonly denoted by the symbol $\text{proj}_{\mathbf{a}}\mathbf{u}$, in which case it follows from (8) that $\mathbf{w}_2 = \mathbf{u} - \text{proj}_{\mathbf{a}}\mathbf{u}$. In summary,

$$\boxed{\text{proj}_{\mathbf{a}}\mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} \quad (\textit{vector component of } \mathbf{u} \textit{ along } \mathbf{a})} \tag{10}$$

$$\boxed{\mathbf{u} - \text{proj}_{\mathbf{a}}\mathbf{u} = \mathbf{u} - \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} \quad (\textit{vector component of } \mathbf{u} \textit{ orthogonal to } \mathbf{a})} \tag{11}$$

EXAMPLE 4 | Vector Component of u Along a

Let $\mathbf{u} = (2, -1, 3)$ and $\mathbf{a} = (4, -1, 2)$. Find the vector component of $\mathbf{u}$ along $\mathbf{a}$ and the vector component of $\mathbf{u}$ orthogonal to $\mathbf{a}$.

Solution

$$\mathbf{u} \cdot \mathbf{a} = (2)(4) + (-1)(-1) + (3)(2) = 15$$

$$\|\mathbf{a}\|^2 = 4^2 + (-1)^2 + 2^2 = 21$$

Thus the vector component of $\mathbf{u}$ along $\mathbf{a}$ is

$$\text{proj}_{\mathbf{a}}\mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} = \tfrac{15}{21}(4, -1, 2) = \left(\tfrac{20}{7}, -\tfrac{5}{7}, \tfrac{10}{7}\right)$$

and the vector component of $\mathbf{u}$ orthogonal to $\mathbf{a}$ is

$$\mathbf{u} - \text{proj}_{\mathbf{a}}\mathbf{u} = (2, -1, 3) - \left(\tfrac{20}{7}, -\tfrac{5}{7}, \tfrac{10}{7}\right) = \left(-\tfrac{6}{7}, -\tfrac{2}{7}, \tfrac{11}{7}\right)$$

As a check, you may wish to verify that the vectors $\mathbf{u} - \text{proj}_{\mathbf{a}}\mathbf{u}$ and $\mathbf{a}$ are perpendicular by showing that their dot product is zero.

EXAMPLE 5 | Orthogonal Projection onto a Line Through the Origin

(*a*) Find the orthogonal projections of the standard unit vectors $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$ onto the line L that makes an angle θ with the positive x-axis.

(*b*) Use the result in part (*a*) to find the standard matrix for the operator $T: R^2 \to R^2$ that maps each point orthogonally onto L.

Solution (*a*) As illustrated in **Figure 3.3.3**, the vector $\mathbf{a} = (\cos\theta, \sin\theta)$ is a unit vector along the line L, so our first problem is to find the orthogonal projection of $\mathbf{e}_1$ along $\mathbf{a}$. Since

$$\|\mathbf{a}\| = \sqrt{\sin^2\theta + \cos^2\theta} = 1 \quad \text{and} \quad \mathbf{e}_1 \cdot \mathbf{a} = (1, 0) \cdot (\cos\theta, \sin\theta) = \cos\theta$$

it follows from Formula (10) that this projection is

$$\text{proj}_{\mathbf{a}}\mathbf{e}_1 = \frac{\mathbf{e}_1 \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} = (\cos\theta)(\cos\theta, \sin\theta) = (\cos^2\theta, \sin\theta\cos\theta)$$

Similarly, since $\mathbf{e}_2 \cdot \mathbf{a} = (0, 1) \cdot (\cos\theta, \sin\theta) = \sin\theta$, it follows from Formula (10) that

$$\text{proj}_{\mathbf{a}}\mathbf{e}_2 = \frac{\mathbf{e}_2 \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} = (\sin\theta)(\cos\theta, \sin\theta) = (\sin\theta\cos\theta, \sin^2\theta)$$

Solution (*b*) It follows from part (*a*) that the standard matrix for T is

$$A = [T(\mathbf{e}_1) \mid T(\mathbf{e}_2)] = \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix} = \begin{bmatrix} \cos^2\theta & \frac{1}{2}\sin 2\theta \\ \frac{1}{2}\sin 2\theta & \sin^2\theta \end{bmatrix}$$

In keeping with common usage, we will denote this matrix by

$$\boxed{P_\theta = \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix} = \begin{bmatrix} \cos^2\theta & \frac{1}{2}\sin 2\theta \\ \frac{1}{2}\sin 2\theta & \sin^2\theta \end{bmatrix}} \tag{12}$$

We have included two versions of Formula (12) because both are commonly used. Whereas the first version involves only the angle θ, the second involves both θ and 2θ.

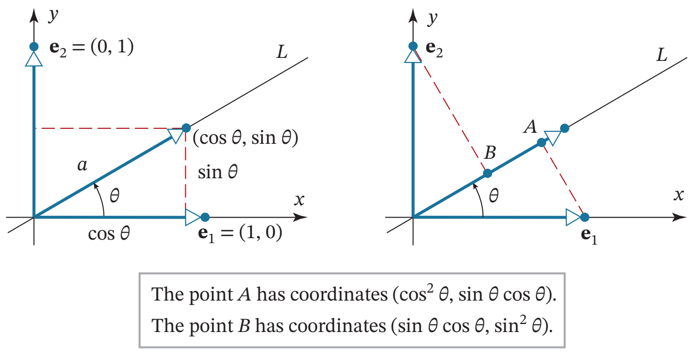


FIGURE 3.3.3

EXAMPLE 6 | Orthogonal Projection onto a Line Through the Origin

Use Formula (12) to find the orthogonal projection of the vector $\mathbf{x} = (1, 5)$ onto the line through the origin that makes an angle of $\pi/6\,(= 30°)$ with the positive x-axis.

Solution Since $\sin(\pi/6) = 1/2$ and $\cos(\pi/6) = \sqrt{3}/2$, it follows from (12) that the standard matrix for this projection is

$$P_{\pi/6} = \begin{bmatrix} \cos^2(\pi/6) & \sin(\pi/6)\cos(\pi/6) \\ \sin(\pi/6)\cos(\pi/6) & \sin^2(\pi/6) \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{1}{4} \end{bmatrix}$$

Thus,

$$P_{\pi/6}\mathbf{x} = \begin{bmatrix} \frac{3}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{1}{4} \end{bmatrix}\begin{bmatrix}1\\5\end{bmatrix} = \begin{bmatrix} \frac{3+5\sqrt{3}}{4} \\ \frac{\sqrt{3}+5}{4} \end{bmatrix} \approx \begin{bmatrix}2.91\\1.68\end{bmatrix}$$

or in comma-delimited notation, $P_{\pi/6}(1, 5) \approx (2.91, 1.68)$.

Reflections About Lines Through the Origin

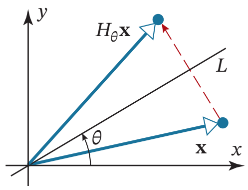


FIGURE 3.3.4

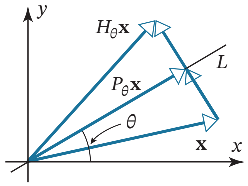


FIGURE 3.3.5

In Table 1 of Section 1.8 we listed the reflections about the coordinate axes in R^2. These are special cases of the more general operator $H_\theta : R^2 \to R^2$ that maps each point into its reflection about a line L through the origin that makes an angle θ with the positive x-axis (**Figure 3.3.4**). We could find the standard matrix for H_θ by finding the images of the standard basis vectors, but instead we will take advantage of our work on orthogonal projections by using Formula (12) for P_θ to find a formula for H_θ.

You should be able to see from **Figure 3.3.5** that for every vector $\mathbf{x}$ in R^n

$$P_\theta\mathbf{x} - \mathbf{x} = \tfrac{1}{2}(H_\theta\mathbf{x} - \mathbf{x}) \quad \text{or equivalently} \quad H_\theta\mathbf{x} = (2P_\theta - I)\mathbf{x}$$

Thus, it follows from Theorem 1.8.4 that

$$H_\theta = 2P_\theta - I \tag{13}$$

and hence from (12) that

$$\boxed{H_\theta = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}} \tag{14}$$

EXAMPLE 7 | Reflection About a Line Through the Origin

Find the reflection of the vector $\mathbf{x} = (1, 5)$ about the line through the origin that makes an angle of $\pi/6$ $(= 30^\circ)$ with the x-axis.

Solution Since $\sin(\pi/3) = \sqrt{3}/2$ and $\cos(\pi/3) = 1/2$, it follows from (14) that the standard matrix for this reflection is

$$H_{\pi/6} = \begin{bmatrix} \cos(\pi/3) & \sin(\pi/3) \\ \sin(\pi/3) & -\cos(\pi/3) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$$

Thus,

$$H_{\pi/6}\mathbf{x} = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix}1\\5\end{bmatrix} = \begin{bmatrix} \frac{1+5\sqrt{3}}{2} \\ \frac{\sqrt{3}-5}{2} \end{bmatrix} \approx \begin{bmatrix}4.83\\-1.63\end{bmatrix}$$

or in comma-delimited notation, $H_{\pi/6}(1, 5) \approx (4.83, -1.63)$.

Norm of a Projection

Sometimes we will be more interested in the *norm* of the vector component of $\mathbf{u}$ along $\mathbf{a}$ than in the vector component itself. A formula for this norm can be derived as follows:

$$\|\text{proj}_{\mathbf{a}}\mathbf{u}\| = \left\|\frac{\mathbf{u}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a}\right\| = \left|\frac{\mathbf{u}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\right|\|\mathbf{a}\| = \frac{|\mathbf{u}\cdot\mathbf{a}|}{\|\mathbf{a}\|^2}\|\mathbf{a}\|$$

where the second equality follows from part (c) of Theorem 3.2.1 and the third from the fact that $\|\mathbf{a}\|^2 > 0$. Thus,

$$\boxed{\|\text{proj}_{\mathbf{a}}\mathbf{u}\| = \frac{|\mathbf{u}\cdot\mathbf{a}|}{\|\mathbf{a}\|}} \tag{15}$$

If θ denotes the angle between $\mathbf{u}$ and $\mathbf{a}$, then $\mathbf{u}\cdot\mathbf{a} = \|\mathbf{u}\|\,\|\mathbf{a}\|\cos\theta$, so (15) can also be written as

$$\boxed{\|\text{proj}_{\mathbf{a}}\mathbf{u}\| = \|\mathbf{u}\|\,|\cos\theta|} \tag{16}$$

(Verify.) A geometric interpretation of this result is given in **Figure 3.3.6**.

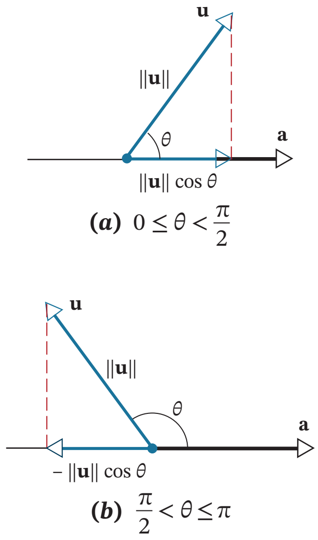


FIGURE 3.3.6

The Theorem of Pythagoras

In Section 3.2 we found that many theorems about vectors in R^2 and R^3 also hold in R^n. Another example of this is the following generalization of the Theorem of Pythagoras (**Figure 3.3.7**).

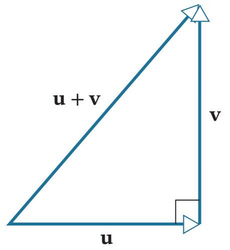


FIGURE 3.3.7

Theorem 3.3.3

Theorem of Pythagoras in R^n

If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal vectors in R^n with the Euclidean inner product, then

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \tag{17}$$

Proof Since $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, we have $\mathbf{u}\cdot\mathbf{v} = 0$, from which it follows that

$$\|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) = \|\mathbf{u}\|^2 + 2(\mathbf{u}\cdot\mathbf{v}) + \|\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \quad \blacksquare$$

EXAMPLE 8 | Theorem of Pythagoras in R^4

We showed in Example 1 that the vectors

$$\mathbf{u} = (-2, 3, 1, 4) \quad \text{and} \quad \mathbf{v} = (1, 2, 0, -1)$$

are orthogonal. Verify the Theorem of Pythagoras for these vectors.

Solution We leave it for you to confirm that

$$\begin{aligned} &\mathbf{u}+\mathbf{v} = (-1, 5, 1, 3) \\ &\|\mathbf{u}+\mathbf{v}\|^2 = 36 \\ &\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = 30 + 6 \end{aligned}$$

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

Distance Problems

OPTIONAL: We will now show how orthogonal projections can be used to solve the following three distance problems:

Problem 1. Find the distance between a point and a line in R^2.

Problem 2. Find the distance between a point and a plane in R^3.

Problem 3. Find the distance between two parallel planes in R^3.

A method for solving the first two problems is provided by the next theorem. Since the proofs of the two parts are similar, we will prove part (*b*) and leave part (*a*) as an exercise.

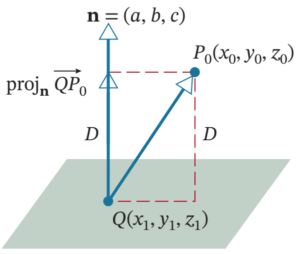
Theorem 3.3.4

(*a*) In R^2 the distance D between the point $P_0(x_0, y_0)$ and the line $ax + by + c = 0$ is

$$D = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}} \tag{18}$$

(*b*) In R^3 the distance D between the point $P_0(x_0, y_0, z_0)$ and the plane $ax + by + cz + d = 0$ is

$$D = \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}} \tag{19}$$



Distance from P_0 to plane.

FIGURE 3.3.8

Proof (b) The underlying idea of the proof is illustrated in **Figure 3.3.8**. As shown in that figure, let $Q(x_1, y_1, z_1)$ be any point in the plane, and let $\mathbf{n} = (a, b, c)$ be a normal vector to the plane that is positioned with its initial point at Q. It is now evident that the distance D between P_0 and the plane is simply the length (or norm) of the orthogonal projection of the vector $\overrightarrow{QP_0}$ on $\mathbf{n}$, which by Formula (15) is

$$D = \|\text{proj}_{\mathbf{n}}\overrightarrow{QP_0}\| = \frac{|\overrightarrow{QP_0} \cdot \mathbf{n}|}{\|\mathbf{n}\|}$$

But

$$\overrightarrow{QP_0} = (x_0 - x_1, y_0 - y_1, z_0 - z_1)$$

$$\overrightarrow{QP_0} \cdot \mathbf{n} = a(x_0 - x_1) + b(y_0 - y_1) + c(z_0 - z_1)$$

$$\|\mathbf{n}\| = \sqrt{a^2 + b^2 + c^2}$$

Thus

$$D = \frac{|a(x_0 - x_1) + b(y_0 - y_1) + c(z_0 - z_1)|}{\sqrt{a^2 + b^2 + c^2}} \tag{20}$$

Since the point $Q(x_1, y_1, z_1)$ lies in the given plane, its coordinates satisfy the equation of that plane; thus

$$ax_1 + by_1 + cz_1 + d = 0$$

or

$$d = -ax_1 - by_1 - cz_1$$

Substituting this expression in (20) yields (19). ■

EXAMPLE 9 | Distance Between a Point and a Plane

Find the distance D between the point $(1, -4, -3)$ and the plane $2x - 3y + 6z = -1$.

Solution Since the distance formulas in Theorem 3.3.4 require that the equations of the line and plane be written with zero on the right side, we first need to rewrite the equation of the plane as

$$2x - 3y + 6z + 1 = 0$$

from which we obtain

$$D = \frac{|2(1) + (-3)(-4) + 6(-3) + 1|}{\sqrt{2^2 + (-3)^2 + 6^2}} = \frac{|-3|}{7} = \frac{3}{7}$$

The third distance problem posed above is to find the distance between two parallel planes in R^3. As suggested in **Figure 3.3.9**, the distance between a plane V and a plane W can be obtained by finding any point P_0 in one of the planes, and computing the distance between that point and the other plane. Here is an example.

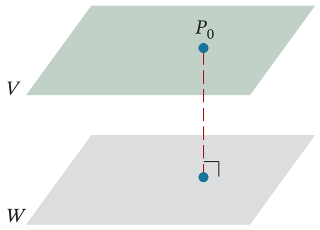


FIGURE 3.3.9 The distance between the parallel planes V and W is equal to the distance between P_0 and W.

EXAMPLE 10 | Distance Between Parallel Planes

The planes

$$x + 2y - 2z = 3 \quad \text{and} \quad 2x + 4y - 4z = 7$$

are parallel since their normals, $(1, 2, -2)$ and $(2, 4, -4)$, are parallel vectors. Find the distance between these planes.

Solution To find the distance D between the planes, we can select an arbitrary point in one of the planes and compute its distance to the other plane. By setting $y = z = 0$ in the equation $x + 2y - 2z = 3$, we obtain the point $P_0(3, 0, 0)$ in this plane. From (19), the distance between P_0 and the plane $2x + 4y - 4z = 7$ is

$$D = \frac{|2(3) + 4(0) + (-4)(0) - 7|}{\sqrt{2^2 + 4^2 + (-4)^2}} = \frac{1}{6}$$

Exercise Set 3.3

In Exercises 1–2, determine whether **u** *and* **v** *are orthogonal vectors.*

1. a. $\mathbf{u} = (6, 1, 4),\ \mathbf{v} = (2, 0, -3)$
   b. $\mathbf{u} = (0, 0, -1),\ \mathbf{v} = (1, 1, 1)$
   c. $\mathbf{u} = (3, -2, 1, 3),\ \mathbf{v} = (-4, 1, -3, 7)$
   d. $\mathbf{u} = (5, -4, 0, 3),\ \mathbf{v} = (-4, 1, -3, 7)$

2. a. $\mathbf{u} = (2, 3),\ \mathbf{v} = (5, -7)$
   b. $\mathbf{u} = (1, 1, 1),\ \mathbf{v} = (0, 0, 0)$
   c. $\mathbf{u} = (1, -5, 4),\ \mathbf{v} = (3, 3, 3)$
   d. $\mathbf{u} = (4, 1, -2, 5),\ \mathbf{v} = (-1, 5, 3, 1)$

In Exercises 3–6, find a point-normal form of the equation of the plane passing through P and having **n** *as a normal.*

3. $P(-1, 3, -2);\ \mathbf{n} = (-2, 1, -1)$
4. $P(1, 1, 4);\ \mathbf{n} = (1, 9, 8)$
5. $P(2, 0, 0);\ \mathbf{n} = (0, 0, 2)$
6. $P(0, 0, 0);\ \mathbf{n} = (1, 2, 3)$

In Exercises 7–10, determine whether the given planes are parallel.

7. $4x - y + 2z = 5$ and $7x - 3y + 4z = 8$
8. $x - 4y - 3z - 2 = 0$ and $3x - 12y - 9z - 7 = 0$
9. $2y = 8x - 4z + 5$ and $x = \frac{1}{2}z + \frac{1}{4}y$
10. $(-4, 1, 2) \cdot (x, y, z) = 0$ and $(8, -2, -4) \cdot (x, y, z) = 0$

In Exercises 11–12, determine whether the given planes are perpendicular.

11. $3x - y + z - 4 = 0,\ x + 2z = -1$
12. $x - 2y + 3z = 4,\ -2x + 5y + 4z = -1$

In Exercises 13–14, find $\|\text{proj}_{\mathbf{a}}\mathbf{u}\|$.

13. a. $\mathbf{u} = (1, -2),\ \mathbf{a} = (-4, -3)$
    b. $\mathbf{u} = (3, 0, 4),\ \mathbf{a} = (2, 3, 3)$

14. a. $\mathbf{u} = (5, 6),\ \mathbf{a} = (2, -1)$
    b. $\mathbf{u} = (3, -2, 6),\ \mathbf{a} = (1, 2, -7)$

In Exercises 15–20, find the vector component of **u** *along* **a** *and the vector component of* **u** *orthogonal to* **a**.

15. $\mathbf{u} = (6, 2),\ \mathbf{a} = (3, -9)$
16. $\mathbf{u} = (-1, -2),\ \mathbf{a} = (-2, 3)$
17. $\mathbf{u} = (3, 1, -7),\ \mathbf{a} = (1, 0, 5)$
18. $\mathbf{u} = (2, 0, 1),\ \mathbf{a} = (1, 2, 3)$
19. $\mathbf{u} = (2, 1, 1, 2),\ \mathbf{a} = (4, -4, 2, -2)$
20. $\mathbf{u} = (5, 0, -3, 7),\ \mathbf{a} = (2, 1, -1, -1)$

In Exercises 21–24, find the distance between the point and the line.

21. $(-3, 1);\ 4x + 3y + 4 = 0$
22. $(-1, 4);\ x - 3y + 2 = 0$
23. $(2, -5);\ y = -4x + 2$
24. $(1, 8);\ 3x + y = 5$

In Exercises 25–26, find the distance between the point and the plane.

25. $(3, 1, -2);\ x + 2y - 2z = 4$
26. $(-1, -1, 2);\ 2x + 5y - 6z = 4$

In Exercises 27–28, find the distance between the given parallel planes.

27. $2x - y - z = 5$ and $-4x + 2y + 2z = 12$

28. $2x - y + z = 1$ and $2x - y + z = -1$

29. Find a unit vector that is orthogonal to both $\mathbf{u} = (1, 0, 1)$ and $\mathbf{v} = (0, 1, 1)$.

30. **a.** Show that $\mathbf{v} = (a, b)$ and $\mathbf{w} = (-b, a)$ are orthogonal vectors.

b. Use the result in part (a) to find two vectors that are orthogonal to $\mathbf{v} = (2, -3)$.

c. Find two unit vectors that are orthogonal to $\mathbf{v} = (-3, 4)$.

31. Do the points $A(1, 1, 1)$, $B(-2, 0, 3)$, and $C(-3, -1, 1)$ form the vertices of a right triangle? Explain.

32. Repeat Exercise 31 for the points $A(3, 0, 2)$, $B(4, 3, 0)$, and $C(8, 1, -1)$.

33. Show that if $\mathbf{v}$ is orthogonal to both $\mathbf{w}_1$ and $\mathbf{w}_2$, then $\mathbf{v}$ is orthogonal to $k_1\mathbf{w}_1 + k_2\mathbf{w}_2$ for all scalars k_1 and k_2.

34. Is it possible to have $\text{proj}_{\mathbf{a}}\mathbf{u} = \text{proj}_{\mathbf{u}}\mathbf{a}$? Explain.

In Exercises 35–36, find the standard matrix for the reflection of R^2 about the stated line, and then use that matrix to find the reflection of the given point about that line.

35. The reflection of $(3, 4)$ about the line that makes an angle of $\pi/3$ $(= 60^\circ)$ with the positive x-axis.

36. The reflection of $(1, 2)$ about the line that makes an angle of $\pi/4$ $(= 45^\circ)$ with the positive x-axis.

In Exercises 37–38, find the standard matrix for the orthogonal projection of R^2 onto the stated line, and then use that matrix to find the orthogonal projection of the given point onto that line.

37. The orthogonal projection of $(3, 4)$ onto the line that makes an angle of $\pi/3$ $(= 60^\circ)$ with the positive x-axis.

38. The orthogonal projection of $(1, 2)$ onto the line that makes an angle of $\pi/4$ $(= 45^\circ)$ with the positive x-axis.

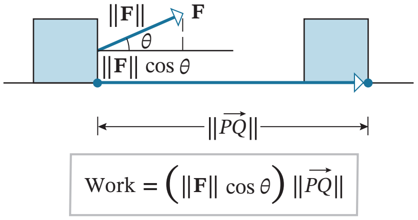
Exercises 39–41 *In physics and engineering the **work** W performed by a* constant force $\mathbf{F}$ *applied in the* direction of motion *to an object moving a distance* d *on a straight line is defined to be*

$$W = \|\mathbf{F}\|d \quad \text{(force magnitude times distance)}$$

*In the case where the applied force is constant but makes an angle θ with the direction of motion, and where the object moves along a line from a point P to a point Q, we call $\overrightarrow{PQ}$ the **displacement** and define the work performed by the force to be*

$$W = \mathbf{F} \cdot \overrightarrow{PQ} = \|\mathbf{F}\|\,\|\overrightarrow{PQ}\| \cos\theta$$

(see accompanying figure). Common units of work are ft-lb (foot pounds) or Nm (Newton meters).

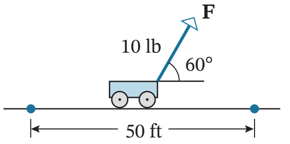


39. Show that the work performed by a constant force (not necessarily in the direction of motion) can be expressed as

$$W = \pm\|\overrightarrow{PQ}\|\,\|\text{proj}_{\overrightarrow{PQ}}\mathbf{F}\|$$

and explain when the + sign should be used and when the − sign should be used.

40. As illustrated in the accompanying figure, a wagon is pulled horizontally by exerting a force of 10 lb on the handle at an angle of 60° with the horizontal. How much work is done in moving the wagon 50 ft?



41. A sailboat travels 100 m due north while the wind exerts a force of 500 N toward the northeast. How much work does the wind do?

Working with Proofs

42. Let $\mathbf{u}$ and $\mathbf{v}$ be nonzero vectors in 2- or 3-space, and let $k = \|\mathbf{u}\|$ and $l = \|\mathbf{v}\|$. Prove that the vector $\mathbf{w} = l\mathbf{u} + k\mathbf{v}$ bisects the angle between $\mathbf{u}$ and $\mathbf{v}$.

43. Prove part (a) of Theorem 3.3.4.

44. In R^3 the ***orthogonal projections*** onto the x-axis, y-axis, and z-axis are

$$T_1(x, y, z) = (x, 0, 0), \quad T_2(x, y, z) = (0, y, 0),$$
$$T_3(x, y, z) = (0, 0, z)$$

respectively.

a. Show that if $T\colon R^3 \to R^3$ is an orthogonal projection onto one of the coordinate axes, then for every vector $\mathbf{x}$ in R^3, the vectors $T(\mathbf{x})$ and $\mathbf{x} - T(\mathbf{x})$ are orthogonal.

b. Make a sketch showing $\mathbf{x}$ and $\mathbf{x} - T(\mathbf{x})$ in the case where T is the orthogonal projection onto the x-axis.

45. **a.** Use Formula (14) and appropriate trigonometric identities to prove that multiplication by the matrix

$$H_m = \frac{1}{1 + m^2}\begin{bmatrix} 1 - m^2 & 2m \\ 2m & m^2 - 1 \end{bmatrix}$$

performs a reflection about the line $y = mx$.

b. Use the result in part (a) to show that multiplication by the matrix

$$H = \begin{bmatrix} \frac{5}{13} & \frac{12}{13} \\ \frac{12}{13} & -\frac{5}{13} \end{bmatrix}$$

performs a reflection about a line through the origin, and find an equation for that line.

True-False Exercises

TF. In parts (a)–(g) determine whether the statement is true or false, and justify your answer.

a. The vectors $(3, -1, 2)$ and $(0, 0, 0)$ are orthogonal.

b. If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal vectors, then for all nonzero scalars k and m, $k\mathbf{u}$ and $m\mathbf{v}$ are orthogonal vectors.

c. The orthogonal projection of $\mathbf{u}$ on $\mathbf{a}$ is perpendicular to the vector component of $\mathbf{u}$ orthogonal to $\mathbf{a}$.

d. If $\mathbf{a}$ and $\mathbf{b}$ are orthogonal vectors, then for every nonzero vector $\mathbf{u}$, we have

$$\text{proj}_{\mathbf{a}}(\text{proj}_{\mathbf{b}}(\mathbf{u})) = \mathbf{0}$$

e. If $\mathbf{a}$ and $\mathbf{u}$ are nonzero vectors, then

$$\text{proj}_{\mathbf{a}}(\text{proj}_{\mathbf{a}}(\mathbf{u})) = \text{proj}_{\mathbf{a}}(\mathbf{u})$$

f. If the relationship

$$\text{proj}_{\mathbf{a}}\mathbf{u} = \text{proj}_{\mathbf{a}}\mathbf{v}$$

holds for some nonzero vector $\mathbf{a}$, then $\mathbf{u} = \mathbf{v}$.

g. For all vectors $\mathbf{u}$ and $\mathbf{v}$, it is true that

$$\|\mathbf{u} + \mathbf{v}\| = \|\mathbf{u}\| + \|\mathbf{v}\|$$

Working with Technology

T1. Find the lengths of the sides and the interior angles of the triangle in R^4 whose vertices are

$$P(2,4,2,4,2), \quad Q(6,4,4,4,6), \quad R(5,7,5,7,2)$$

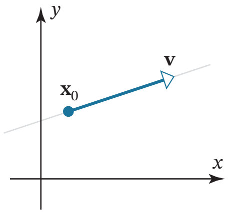
T2. Express the vector $\mathbf{u} = (2,3,1,2)$ in the form $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1$ is a scalar multiple of $\mathbf{a} = (-1,0,2,1)$ and $\mathbf{w}_2$ is orthogonal to $\mathbf{a}$.

3.4 The Geometry of Linear Systems

In this section we will use parametric and vector methods to study general systems of linear equations. This work will enable us to interpret solution sets of linear systems with n unknowns as geometric objects in R^n just as we interpreted solution sets of linear systems with two and three unknowns as points, lines, and planes in R^2 and R^3.

Vector and Parametric Equations of Lines in R^2 and R^3

In the last section we derived equations of lines and planes that are determined by a point and a normal vector. However, there are other useful ways of specifying lines and planes. For example, a unique line in R^2 or R^3 is determined by a point $\mathbf{x}_0$ on the line and a nonzero vector $\mathbf{v}$ parallel to the line, and a unique plane in R^3 is determined by a point $\mathbf{x}_0$ in the plane and two noncollinear vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ parallel to the plane. The best way to visualize the latter is to translate the vectors so their initial points are at $\mathbf{x}_0$ (**Figure 3.4.1**).



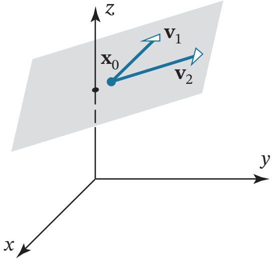


FIGURE 3.4.1

Let us begin by deriving an equation for the line L that contains a point $\mathbf{x}_0$ and is parallel to a nonzero vector $\mathbf{v}$. If $\mathbf{x}$ is a general point on such a line, then, as illustrated in **Figure 3.4.2**, the vector $\mathbf{x} - \mathbf{x}_0$ will be some scalar multiple of $\mathbf{v}$, say

$$\mathbf{x} - \mathbf{x}_0 = t\mathbf{v} \quad \text{or equivalently} \quad \mathbf{x} = \mathbf{x}_0 + t\mathbf{v}$$

As the variable t (called a ***parameter***) varies from $-\infty$ to ∞, the point $\mathbf{x}$ traces out the line L. Accordingly, we have the following result.

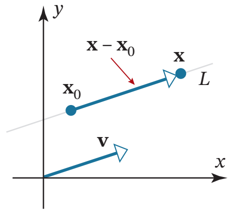


FIGURE 3.4.2

Theorem 3.4.1

Let L be the line in R^2 or R^3 that contains the point $\mathbf{x}_0$ and is parallel to the nonzero vector $\mathbf{v}$. Then the equation of the line through $\mathbf{x}_0$ that is parallel to $\mathbf{v}$ is

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{v} \tag{1}$$

If $\mathbf{x}_0 = \mathbf{0}$, then the line passes through the origin and the equation has the form

$$\mathbf{x} = t\mathbf{v} \tag{2}$$

Although it is not stated explicitly, it is understood in Formulas (1) and (2) that the parameter t varies from $-\infty$ to ∞. This applies to all vector and parametric equations in this text except where stated otherwise.

Vector and Parametric Equations of Planes in R^3

Next we will derive an equation for the plane W that contains a point $\mathbf{x}_0$ and is parallel to the noncollinear vectors $\mathbf{v}_1$ and $\mathbf{v}_2$. As shown in **Figure 3.4.3**, if $\mathbf{x}$ is any point in the plane, then by forming suitable scalar multiples of $\mathbf{v}_1$ and $\mathbf{v}_2$, say $t_1\mathbf{v}_1$ and $t_2\mathbf{v}_2$, we can create a parallelogram with diagonal $\mathbf{x} - \mathbf{x}_0$ and adjacent sides $t_1\mathbf{v}_1$ and $t_2\mathbf{v}_2$. Thus, we have

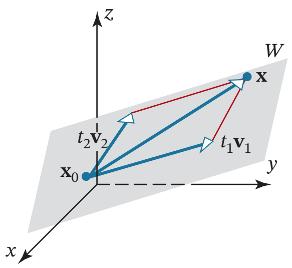


FIGURE 3.4.3

$$\mathbf{x} - \mathbf{x}_0 = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 \quad \text{or equivalently} \quad \mathbf{x} = \mathbf{x}_0 + t_1\mathbf{v}_1 + t_2\mathbf{v}_2$$

As the parameters t_1 and t_2 vary independently from $-\infty$ to ∞, the point $\mathbf{x}$ varies over the entire plane W. In summary, we have the following result.

Theorem 3.4.2

Let W be the plane in R^3 that contains the point $\mathbf{x}_0$ and is parallel to the noncollinear vectors $\mathbf{v}_1$ and $\mathbf{v}_2$. Then an equation of the plane through $\mathbf{x}_0$ that is parallel to $\mathbf{v}_1$ and $\mathbf{v}_2$ is given by

$$\mathbf{x} = \mathbf{x}_0 + t_1\mathbf{v}_1 + t_2\mathbf{v}_2 \tag{3}$$

If $\mathbf{x}_0 = \mathbf{0}$, then the plane passes through the origin and the equation has the form

$$\mathbf{x} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 \tag{4}$$

Remark Observe that the line through $\mathbf{x}_0$ represented by Equation (1) is the translation by $\mathbf{x}_0$ of the line through the origin represented by Equation (2) and that the plane through $\mathbf{x}_0$ represented by Equation (3) is the translation by $\mathbf{x}_0$ of the plane through the origin represented by Equation (4) (**Figure 3.4.4**).

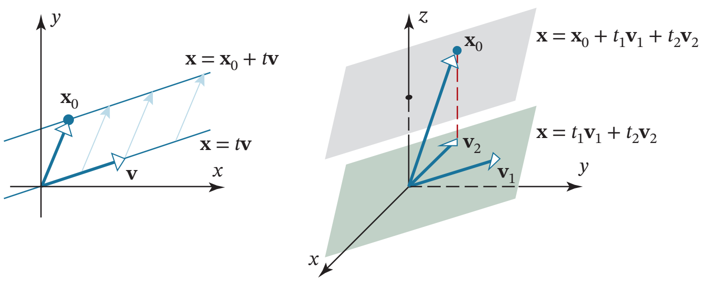


FIGURE 3.4.4

Motivated by the forms of Formulas (1) to (4), we can extend the notions of line and plane to R^n by making the following definitions.

Definition 1

If $\mathbf{x}_0$ and $\mathbf{v}$ are vectors in R^n, and if $\mathbf{v}$ is nonzero, then the equation

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{v} \tag{5}$$

defines the ***line through*** $\mathbf{x}_0$ ***that is parallel to*** $\mathbf{v}$.

Definition 2

If $\mathbf{x}_0$, $\mathbf{v}_1$, and $\mathbf{v}_2$ are nonzero vectors in R^n, and if $\mathbf{v}_1$ and $\mathbf{v}_2$ are not collinear, then the equation

$$\mathbf{x} = \mathbf{x}_0 + t_1\mathbf{v}_1 + t_2\mathbf{v}_2 \tag{6}$$

defines the ***plane through*** $\mathbf{x}_0$ ***that is parallel to*** $\mathbf{v}_1$ ***and*** $\mathbf{v}_2$.

Equations (5) and (6) are called ***vector forms*** of a line and plane in R^n. If the vectors in these equations are expressed in terms of their components and the corresponding components on each side are equated, then the resulting equations are called ***parametric equations*** of the line and plane. Here are some examples.

EXAMPLE 1 | Vector and Parametric Equations of Lines in R^2 and R^3

(*a*) Find a vector equation and parametric equations of the line in R^2 that passes through the origin and is parallel to the vector $\mathbf{v} = (-2, 3)$.

(*b*) Find a vector equation and parametric equations of the line in R^3 that passes through the point $P_0(1, 2, -3)$ and is parallel to the vector $\mathbf{v} = (4, -5, 1)$.

(*c*) Use the vector equation obtained in part (b) to find two points on the line that are different from P_0.

Solution (*a*) It follows from (5) with $\mathbf{x}_0 = \mathbf{0}$ that a vector equation of the line is $\mathbf{x} = t\mathbf{v}$. If we let $\mathbf{x} = (x, y)$, then this equation can be expressed in vector form as

$$(x, y) = t(-2, 3)$$

Equating corresponding components on the two sides of this equation yields the parametric equations

$$x = -2t, \quad y = 3t$$

Solution (*b*) It follows from (5) that a vector equation of the line is $\mathbf{x} = \mathbf{x}_0 + t\mathbf{v}$. If we let $\mathbf{x} = (x, y, z)$, and if we take $\mathbf{x}_0 = (1, 2, -3)$, then this equation can be expressed in vector form as

$$(x, y, z) = (1, 2, -3) + t(4, -5, 1) \tag{7}$$

Equating corresponding components on the two sides of this equation yields the parametric equations

$$x = 1 + 4t, \quad y = 2 - 5t, \quad z = -3 + t$$

Solution (*c*) A point on the line represented by Equation (7) can be obtained by substituting a numerical value for the parameter t. However, since $t = 0$ produces $(x, y, z) = (1, 2, -3)$, which is the point P_0, this value of t does not serve our purpose. Taking $t = 1$ produces the point $(5, -3, -2)$ and taking $t = -1$ produces the point $(-3, 7, -4)$. Any other distinct values for t (except $t = 0$) would work just as well.

EXAMPLE 2 | Vector and Parametric Equations of a Plane in R^3

Find vector and parametric equations of the plane $x - y + 2z = 5$.

Solution We will find the parametric equations first. We can do this by solving the equation for any one of the variables in terms of the other two and then using those two variables as parameters. For example, solving for x in terms of y and z yields

$$x = 5 + y - 2z \tag{8}$$

and then using y and z as parameters t_1 and t_2, respectively, yields the parametric equations

$$x = 5 + t_1 - 2t_2, \quad y = t_1, \quad z = t_2$$

To obtain a vector equation of the plane we rewrite these parametric equations as

$$(x, y, z) = (5 + t_1 - 2t_2, t_1, t_2)$$

or, equivalently, as

$$(x, y, z) = (5, 0, 0) + t_1(1, 1, 0) + t_2(-2, 0, 1)$$

We would have obtained different parametric and vector equations in Example 2 had we solved (8) for y or z rather than x. However, one can show the same plane results in all three cases as the parameters vary from $-\infty$ to ∞.

EXAMPLE 3 | Vector and Parametric Equations of Lines and Planes in R^4

(*a*) Find vector and parametric equations of the line through the origin of R^4 that is parallel to the vector $\mathbf{v} = (5, -3, 6, 1)$.

(*b*) Find vector and parametric equations of the plane in R^4 that passes through the point $\mathbf{x}_0 = (2, -1, 0, 3)$ and is parallel to both $\mathbf{v}_1 = (1, 5, 2, -4)$ and $\mathbf{v}_2 = (0, 7, -8, 6)$.

Solution (*a*) If we let $\mathbf{x} = (x_1, x_2, x_3, x_4)$, then the vector equation $\mathbf{x} = t\mathbf{v}$ can be expressed as

$$(x_1, x_2, x_3, x_4) = t(5, -3, 6, 1)$$

Equating corresponding components yields the parametric equations

$$x_1 = 5t, \quad x_2 = -3t, \quad x_3 = 6t, \quad x_4 = t$$

Solution (*b*) The vector equation $\mathbf{x} = \mathbf{x}_0 + t_1\mathbf{v}_1 + t_2\mathbf{v}_2$ can be expressed as

$$(x_1, x_2, x_3, x_4) = (2, -1, 0, 3) + t_1(1, 5, 2, -4) + t_2(0, 7, -8, 6)$$

which yields the parametric equations

$$\begin{aligned} x_1 &= 2 + t_1 \\ x_2 &= -1 + 5t_1 + 7t_2 \\ x_3 &= 2t_1 - 8t_2 \\ x_4 &= 3 - 4t_1 + 6t_2 \end{aligned}$$

Lines Through Two Points in R^n

If $\mathbf{x}_0$ and $\mathbf{x}_1$ are distinct points in R^n, then the line containing these points is parallel to the vector $\mathbf{v} = \mathbf{x}_1 - \mathbf{x}_0$ (**Figure 3.4.5**), so it follows from (5) that the line can be expressed in vector form as

$$\boxed{\mathbf{x} = \mathbf{x}_0 + t(\mathbf{x}_1 - \mathbf{x}_0)} \tag{9}$$

or, equivalently, as

$$\boxed{\mathbf{x} = (1 - t)\mathbf{x}_0 + t\mathbf{x}_1} \tag{10}$$

These are called the ***two-point vector equations*** of a line in R^n.

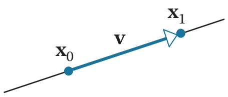


FIGURE 3.4.5

EXAMPLE 4 | A Line Through Two Points in R^2

Find vector and parametric equations for the line in R^2 that passes through the points $P(0,7)$ and $Q(5,0)$.

Solution It does not matter which point we take to be $\mathbf{x}_0$ and which we take to be $\mathbf{x}_1$, so let us arbitrarily choose $\mathbf{x}_0 = (0,7)$ and $\mathbf{x}_1 = (5,0)$. It follows that $\mathbf{x}_1 - \mathbf{x}_0 = (5,-7)$ and hence that

$$(x,y) = (0,7) + t(5,-7) \tag{11}$$

which we can rewrite in parametric form as

$$x = 5t, \quad y = 7 - 7t$$

Had we reversed our choices and taken $\mathbf{x}_0 = (5,0)$ and $\mathbf{x}_1 = (0,7)$, then the resulting vector equation would have been

$$(x,y) = (5,0) + t(-5,7) \tag{12}$$

and the parametric equations would have been

$$x = 5 - 5t, \quad y = 7t$$

(verify). Although (11) and (12) look different, they both represent the line whose equation in rectangular coordinates is

$$7x + 5y = 35$$

(**Figure 3.4.6**). This can be seen by eliminating the parameter t from the parametric equations (verify).

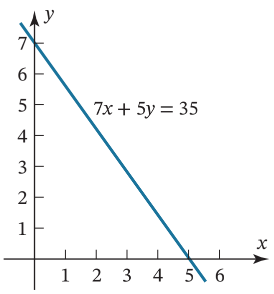


FIGURE 3.4.6

The point $\mathbf{x} = (x,y)$ in Equations (9) and (10) traces an entire line in R^2 as the parameter t varies over the interval $(-\infty, \infty)$. If, however, we restrict the parameter to vary from $t = 0$ to $t = 1$, then $\mathbf{x}$ will not trace the entire line but rather just the *line segment* joining the points $\mathbf{x}_0$ and $\mathbf{x}_1$. The point $\mathbf{x}$ will start at $\mathbf{x}_0$ when $t = 0$ and end at $\mathbf{x}_1$ when $t = 1$. Accordingly, we make the following definition.

Definition 3

If $\mathbf{x}_0$ and $\mathbf{x}_1$ are vectors in R^n, then the equation

$$\mathbf{x} = \mathbf{x}_0 + t(\mathbf{x}_1 - \mathbf{x}_0) \quad (0 \le t \le 1) \tag{13}$$

defines the ***line segment from*** $\mathbf{x}_0$ ***to*** $\mathbf{x}_1$. When convenient, Equation (13) can be written as

$$\mathbf{x} = (1-t)\mathbf{x}_0 + t\mathbf{x}_1 \quad (0 \le t \le 1) \tag{14}$$

EXAMPLE 5 | A Line Segment from One Point to Another in R^2

It follows from (13) and (14) that the line segment in R^2 from $\mathbf{x}_0 = (1,-3)$ to $\mathbf{x}_1 = (5,6)$ can be represented either by the equation

$$\mathbf{x} = (1,-3) + t(4,9) \quad (0 \le t \le 1)$$

or by the equation

$$\mathbf{x} = (1-t)(1,-3) + t(5,6) \quad (0 \le t \le 1)$$

Dot Product Form of a Linear System

Our next objective is to show how to express linear equations and linear systems in dot product notation. This will lead us to some important results about orthogonality and linear systems.

Recall that a *linear equation* in the variables $x_1, x_2, \dots, x_n$ has the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b \quad (a_1, a_2, \dots, a_n \text{ not all zero}) \tag{15}$$

and that the corresponding *homogeneous* equation is

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0 \quad (a_1, a_2, \dots, a_n \text{ not all zero}) \tag{16}$$

These equations can be rewritten in vector form by letting

$$\mathbf{a} = (a_1, a_2, \dots, a_n) \quad \text{and} \quad \mathbf{x} = (x_1, x_2, \dots, x_n)$$

in which case Formula (15) can be written as

$$\boxed{\mathbf{a} \cdot \mathbf{x} = b} \tag{17}$$

and Formula (16) as

$$\boxed{\mathbf{a} \cdot \mathbf{x} = 0} \tag{18}$$

Except for a notational change from $\mathbf{n}$ to $\mathbf{a}$, Formula (18) is the extension to R^n of Formula (6) in Section 3.3. This equation reveals that *each solution vector* $\mathbf{x}$ *of a homogeneous equation is orthogonal to the coefficient vector* $\mathbf{a}$. To take this geometric observation a step further, consider the homogeneous system

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

If we denote the successive row vectors of the coefficient matrix by $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_m$, then we can rewrite this system in dot product form as

$$\begin{array}{c} \mathbf{r}_1 \cdot \mathbf{x} = 0 \\ \mathbf{r}_2 \cdot \mathbf{x} = 0 \\ \vdots \quad \vdots \\ \mathbf{r}_m \cdot \mathbf{x} = 0 \end{array} \tag{19}$$

from which we see that every solution vector $\mathbf{x}$ is orthogonal to every row vector of the coefficient matrix. In summary, we have the following result.

Theorem 3.4.3

If A is an $m \times n$ matrix, then the solution set of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$ consists of all vectors in R^n that are orthogonal to every row vector of A.

EXAMPLE 6 | Orthogonality of Row Vectors and Solution Vectors

We showed in Example 6 of Section 1.2 that the general solution of the homogeneous linear system

$$\begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 \\ 0 & 0 & 5 & 10 & 0 & 15 \\ 2 & 6 & 0 & 8 & 4 & 18 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

is

$$x_1 = -3r - 4s - 2t, \quad x_2 = r, \quad x_3 = -2s, \quad x_4 = s, \quad x_5 = t, \quad x_6 = 0$$

which we can rewrite in vector form as

$$\mathbf{x} = (-3r - 4s - 2t, r, -2s, s, t, 0)$$

According to Theorem 3.4.3, the vector $\mathbf{x}$ must be orthogonal to each of the row vectors

$$\begin{aligned} \mathbf{r}_1 &= (1, 3, -2, 0, 2, 0) \\ \mathbf{r}_2 &= (2, 6, -5, -2, 4, -3) \\ \mathbf{r}_3 &= (0, 0, 5, 10, 0, 15) \\ \mathbf{r}_4 &= (2, 6, 0, 8, 4, 18) \end{aligned}$$

We will confirm that $\mathbf{x}$ is orthogonal to $\mathbf{r}_1$, and leave it for you to verify that $\mathbf{x}$ is orthogonal to the other three row vectors as well. The dot product of $\mathbf{r}_1$ and $\mathbf{x}$ is

$$\mathbf{r}_1 \cdot \mathbf{x} = 1(-3r - 4s - 2t) + 3(r) + (-2)(-2s) + 0(s) + 2(t) + 0(0) = 0$$

which establishes the orthogonality.

Exercise Set 3.4

In Exercises 1–4, find vector and parametric equations of the line containing the point and parallel to the vector.

1. Point: $(-4, 1)$; vector: $\mathbf{v} = (0, -8)$

2. Point: $(2, -1)$; vector: $\mathbf{v} = (-4, -2)$

3. Point: $(0, 0, 0)$; vector: $\mathbf{v} = (-3, 0, 1)$

4. Point: $(-9, 3, 4)$; vector: $\mathbf{v} = (-1, 6, 0)$

In Exercises 5–8, use the given equation of a line to find a point on the line and a vector parallel to the line.

5. $\mathbf{x} = (3 - 5t, -6 - t)$

6. $(x, y, z) = (4t, 7, 4 + 3t)$

7. $\mathbf{x} = (1 - t)(4, 6) + t(-2, 0)$

8. $\mathbf{x} = (1 - t)(0, -5, 1)$

In Exercises 9–12, find vector and parametric equations of the plane that contains the given point and is parallel to the two vectors.

9. Point: $(-3, 1, 0)$; vectors: $\mathbf{v}_1 = (0, -3, 6)$ and $\mathbf{v}_2 = (-5, 1, 2)$

10. Point: $(0, 6, -2)$; vectors: $\mathbf{v}_1 = (0, 9, -1)$ and $\mathbf{v}_2 = (0, -3, 0)$

11. Point: $(-1, 1, 4)$; vectors: $\mathbf{v}_1 = (6, -1, 0)$ and $\mathbf{v}_2 = (-1, 3, 1)$

12. Point: $(0, 5, -4)$; vectors: $\mathbf{v}_1 = (0, 0, -5)$ and $\mathbf{v}_2 = (1, -3, -2)$

In Exercises 13–14, find vector and parametric equations of the line in R^2 that passes through the origin and is orthogonal to $\mathbf{v}$.

13. $\mathbf{v} = (-2, 3)$

14. $\mathbf{v} = (1, -4)$

In Exercises 15–16, find vector and parametric equations of the plane in R^3 that passes through the origin and is orthogonal to $\mathbf{v}$.

15. $\mathbf{v} = (4, 0, -5)$ [*Hint:* Construct two nonparallel vectors orthogonal to $\mathbf{v}$ in R^3.]

16. $\mathbf{v} = (3, 1, -6)$

In Exercises 17–20, find the general solution to the linear system and confirm that the row vectors of the coefficient matrix are orthogonal to the solution vectors.

17. $$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ 2x_1 + 2x_2 + 2x_3 &= 0 \\ 3x_1 + 3x_2 + 3x_3 &= 0 \end{aligned}$$

18. $$\begin{aligned} x_1 + 3x_2 - 4x_3 &= 0 \\ 2x_1 + 6x_2 - 8x_3 &= 0 \end{aligned}$$

19. $x_1 + 5x_2 + x_3 + 2x_4 - x_5 = 0$
$x_1 - 2x_2 - x_3 + 3x_4 + 2x_5 = 0$

20. $x_1 + 3x_2 - 4x_3 = 0$
$x_1 + 2x_2 + 3x_3 = 0$

21. a. Find a homogeneous linear system of two equations in three unknowns whose solution space consists of those vectors in R^3 that are orthogonal to the vectors $\mathbf{a} = (1, 1, 1)$ and $\mathbf{b} = (-2, 3, 0)$.

b. What kind of geometric object is the solution space?

c. Find a general solution of the system obtained in part (a), and confirm that Theorem 3.4.3 holds.

22. a. Find a homogeneous linear system of two equations in three unknowns whose solution space consists of those vectors in R^3 that are orthogonal to $\mathbf{a} = (-3, 2, -1)$ and $\mathbf{b} = (0, -2, -2)$.

b. What kind of geometric object is the solution space?

c. Find a general solution of the system obtained in part (a), and confirm that Theorem 3.4.3 holds.

23. a. Let $\mathbf{x} = \mathbf{x}_0 + t\mathbf{v}$ be a line in R^n and let $T_A\colon R^n \to R^n$ be an invertible matrix operator on R^n. Show that the image of a line under multiplication by A is itself a line.

b. Let $T_A\colon R^2 \to R^2$ be multiplication by the matrix

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

Find vector and parametric equations for the image under multiplication by A of the line $\mathbf{x} = (1, 3) + t(2, -1)$.

24. Let $T_A\colon R^3 \to R^3$ be multiplication by the matrix

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

Find a vector equation for the image under multiplication by A of the line segment

$$(x,\ y,\ z) = (1 - t)(2,\ -3,\ 1) + t(4,\ 1,\ 2) \quad (0 \le t \le 1)$$

True-False Exercises

TF. In parts **(a)–(e)** determine whether the statement is true or false, and justify your answer.

a. The vector equation of a line can be determined from any point lying on the line and a nonzero vector parallel to the line.

b. The vector equation of a plane can be determined from any point lying in the plane and a nonzero vector parallel to the plane.

c. The points lying on a line through the origin in R^2 or R^3 are all scalar multiples of any nonzero vector on the line.

d. All solution vectors of the linear system $A\mathbf{x} = \mathbf{b}$ are orthogonal to the row vectors of the matrix A if and only if $\mathbf{b} = \mathbf{0}$.

e. If $\mathbf{x}_1$ and $\mathbf{x}_2$ are two solutions of the nonhomogeneous linear system $A\mathbf{x} = \mathbf{b}$, then $\mathbf{x}_1 - \mathbf{x}_2$ is a solution of the corresponding homogeneous linear system.

Working with Technology

T1. Find the general solution of the homogeneous linear system

$$\begin{bmatrix} 2 & 6 & -4 & 0 & 4 & 0 \\ 0 & 0 & 1 & 2 & 0 & 3 \\ 6 & 18 & -15 & -6 & 12 & -9 \\ 1 & 3 & 0 & 4 & 2 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and confirm that each solution vector is orthogonal to every row vector of the coefficient matrix in accordance with Theorem 3.4.3.

3.5 Cross Product

This optional section is concerned with properties of vectors in 3-space that are important to physicists and engineers. It can be omitted, if desired, since subsequent sections do not depend on its content. Among other things, we define an operation that provides a way of constructing a vector in 3-space that is perpendicular to two given vectors, and we give a geometric interpretation of 3×3 determinants.

Cross Product of Vectors

In Section 3.2 we defined the dot product of two vectors $\mathbf{u}$ and $\mathbf{v}$ in n-space. That operation produced a *scalar* as its result. We will now define a type of vector multiplication that produces a *vector* as the result but which is applicable only to vectors in 3-space.

Definition 1

If $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ are vectors in 3-space, then the ***cross product*** $\mathbf{u} \times \mathbf{v}$ is the vector defined by

$$\mathbf{u} \times \mathbf{v} = (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1)$$

or, in determinant notation,

$$\mathbf{u} \times \mathbf{v} = \left(\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix}, -\begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix}, \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \right) \tag{1}$$

Remark Instead of memorizing (1), you can obtain the components of $\mathbf{u} \times \mathbf{v}$ as follows:

- Form the 2×3 matrix $\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix}$ whose first row contains the components of $\mathbf{u}$ and whose second row contains the components of $\mathbf{v}$.
- To find the first component of $\mathbf{u} \times \mathbf{v}$, delete the first column and take the determinant; to find the second component, delete the second column and take the negative of the determinant; and to find the third component, delete the third column and take the determinant.

EXAMPLE 1 | Calculating a Cross Product

Find $\mathbf{u} \times \mathbf{v}$, where $\mathbf{u} = (1, 2, -2)$ and $\mathbf{v} = (3, 0, 1)$.

Solution From either (1) or the mnemonic in the preceding remark, we have

$$\begin{aligned} \mathbf{u} \times \mathbf{v} &= \left(\begin{vmatrix} 2 & -2 \\ 0 & 1 \end{vmatrix}, -\begin{vmatrix} 1 & -2 \\ 3 & 1 \end{vmatrix}, \begin{vmatrix} 1 & 2 \\ 3 & 0 \end{vmatrix} \right) \\ &= (2, -7, -6) \end{aligned}$$

The following theorem gives some important relationships between the dot product and cross product and also shows that $\mathbf{u} \times \mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.

Theorem 3.5.1

Relationships Involving Cross Product and Dot Product

If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in 3-space, then

(*a*) $\mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) = 0$ [$\mathbf{u} \times \mathbf{v}$ ***is orthogonal to*** $\mathbf{u}$]

(*b*) $\mathbf{v} \cdot (\mathbf{u} \times \mathbf{v}) = 0$ [$\mathbf{u} \times \mathbf{v}$ ***is orthogonal to*** $\mathbf{v}$]

(*c*) $\|\mathbf{u} \times \mathbf{v}\|^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2$ [***Lagrange's identity***]

(*d*) $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w}$ [***vector triple product***]

(*e*) $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{v} \cdot \mathbf{w})\mathbf{u}$ [***vector triple product***]

The formulas for the vector triple products in parts (*d*) and (*e*) of Theorem 3.5.1 are useful because they allow us to use dot products and scalar multiplications to perform calculations that would otherwise require determinants to calculate the required cross products.

Historical Note

The cross product notation $A \times B$ was introduced by the American physicist and mathematician J. Willard Gibbs, (see p. 163) in a series of unpublished lecture notes for his students at Yale University. It appeared in a published work for the first time in the second edition of the book *Vector Analysis*, by Edwin Wilson (1879–1964), a student of Gibbs. Gibbs originally referred to $A \times B$ as the "skew product."

Proof (a) Let $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$. Then

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

Proof (b) Similar to (a).

Proof (c) Since

$$\|\mathbf{u}\times\mathbf{v}\|^2 = (u_2v_3 - u_3v_2)^2 + (u_3v_1 - u_1v_3)^2 + (u_1v_2 - u_2v_1)^2 \tag{2}$$

and

$$\|\mathbf{u}\|^2\|\mathbf{v}\|^2 - (\mathbf{u}\cdot\mathbf{v})^2 = (u_1^2 + u_2^2 + u_3^2)(v_1^2 + v_2^2 + v_3^2) - (u_1v_1 + u_2v_2 + u_3v_3)^2 \tag{3}$$

the proof can be completed by "multiplying out" the right sides of (2) and (3) and verifying their equality.

Proof (d) and (e) See Exercises 40 and 41 (page 199). ■

EXAMPLE 2 | $\mathbf{u}\times\mathbf{v}$ Is Perpendicular to $\mathbf{u}$ and to $\mathbf{v}$

Consider the vectors

$$\mathbf{u} = (1, 2, -2) \quad \text{and} \quad \mathbf{v} = (3, 0, 1)$$

In Example 1 we showed that

$$\mathbf{u}\times\mathbf{v} = (2, -7, -6)$$

Since

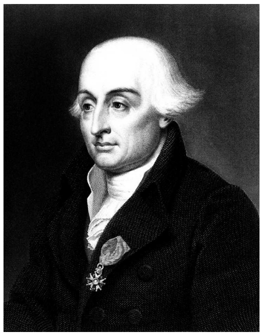
$$\mathbf{u}\cdot(\mathbf{u}\times\mathbf{v}) = (1)(2) + (2)(-7) + (-2)(-6) = 0$$

and

$$\mathbf{v}\cdot(\mathbf{u}\times\mathbf{v}) = (3)(2) + (0)(-7) + (1)(-6) = 0$$

$\mathbf{u}\times\mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$, as guaranteed by Theorem 3.5.1.

Historical Note



Joseph Louis Lagrange (1736–1813)

Joseph Louis Lagrange, who is credited with two of the formulas in Theorem 3.5.1, was a French-Italian mathematician and astronomer. Although his father wanted him to become a lawyer, Lagrange was attracted to mathematics and astronomy after reading a memoir by the astronomer Edmond Halley. At age 16 he began to study mathematics on his own and by age 19 was appointed to a professorship at the Royal Artillery School in Turin. The following year he solved some famous problems using new methods that eventually blossomed into a branch of mathematics called the *calculus of variations*. These methods and Lagrange's applications of them to problems in celestial mechanics were so monumental that by age 25 he was regarded by many of his contemporaries as the greatest living mathematician. One of Lagrange's most famous works is a memoir, *Mécanique Analytique*, in which he reduced the theory of mechanics to a few general formulas from which all other necessary equations could be derived. Napoleon Bonaparte was a great admirer of Lagrange and showered him with many honors. In spite of his fame, Lagrange was a shy and modest man. On his death, he was buried with honor in the Pantheon.

EXAMPLE 3 | Cross Products of the Standard Unit Vectors

Recall from Section 3.2 that the standard unit vectors in 3-space are

$$\mathbf{i} = (1, 0, 0), \quad \mathbf{j} = (0, 1, 0), \quad \mathbf{k} = (0, 0, 1)$$

These vectors each have length 1 and lie along the coordinate axes (**Figure 3.5.1**). Every vector $\mathbf{v} = (v_1, v_2, v_3)$ in 3-space is expressible in terms of $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ since we can write

$$\mathbf{v} = (v_1, v_2, v_3) = v_1(1, 0, 0) + v_2(0, 1, 0) + v_3(0, 0, 1) = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$$

For example,

$$(2, -3, 4) = 2\mathbf{i} - 3\mathbf{j} + 4\mathbf{k}$$

From (1) we obtain

$$\mathbf{i} \times \mathbf{j} = \left(\begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix}, -\begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix}, \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \right) = (0, 0, 1) = \mathbf{k}$$

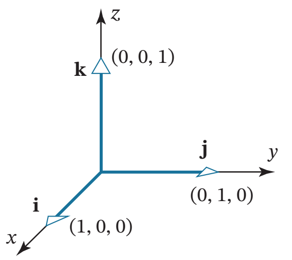


FIGURE 3.5.1 The standard unit vectors.

The main arithmetic properties of the cross product are listed in the next theorem.

Theorem 3.5.2

Properties of Cross Product

If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are any vectors in 3-space and k is any scalar, then:

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

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

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

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

(*e*) $\mathbf{u} \times \mathbf{0} = \mathbf{0} \times \mathbf{u} = \mathbf{0}$

(*f*) $\mathbf{u} \times \mathbf{u} = \mathbf{0}$

The proofs follow immediately from Formula (1) and properties of determinants; for example, part (*a*) can be proved as follows.

Proof (a) Interchanging $\mathbf{u}$ and $\mathbf{v}$ in (1) interchanges the rows of the three determinants on the right side of (1) and hence changes the sign of each component in the cross product. Thus $\mathbf{u} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{u})$. ■

The proofs of the remaining parts are left as exercises.

You should have no trouble obtaining the following results:

$$\begin{array}{lll} \mathbf{i} \times \mathbf{i} = \mathbf{0} & \mathbf{j} \times \mathbf{j} = \mathbf{0} & \mathbf{k} \times \mathbf{k} = \mathbf{0} \\ \mathbf{i} \times \mathbf{j} = \mathbf{k} & \mathbf{j} \times \mathbf{k} = \mathbf{i} & \mathbf{k} \times \mathbf{i} = \mathbf{j} \\ \mathbf{j} \times \mathbf{i} = -\mathbf{k} & \mathbf{k} \times \mathbf{j} = -\mathbf{i} & \mathbf{i} \times \mathbf{k} = -\mathbf{j} \end{array}$$

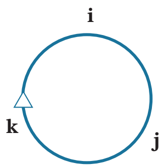


FIGURE 3.5.2

Figure 3.5.2 is helpful for remembering these results. Referring to this diagram, the cross product of two consecutive vectors going clockwise is the next vector around, and the cross product of two consecutive vectors going counterclockwise is the negative of the next vector around.

Determinant Form of Cross Product

It is also worth noting that a cross product can be represented symbolically in the form

$$\boxed{\mathbf{u}\times\mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix}\mathbf{i} - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix}\mathbf{j} + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}\mathbf{k}} \tag{4}$$

For example, if $\mathbf{u} = (1, 2, -2)$ and $\mathbf{v} = (3, 0, 1)$, then

$$\mathbf{u}\times\mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2 & -2 \\ 3 & 0 & 1 \end{vmatrix} = 2\mathbf{i} - 7\mathbf{j} - 6\mathbf{k}$$

which agrees with the result obtained in Example 1.

Remark As evidenced by parts (*d*) and (*e*) of Theorem 3.5.1, it is not true in general that $\mathbf{u}\times(\mathbf{v}\times\mathbf{w}) = (\mathbf{u}\times\mathbf{v})\times\mathbf{w}$. For example,

$$\mathbf{i}\times(\mathbf{j}\times\mathbf{j}) = \mathbf{i}\times\mathbf{0} = \mathbf{0}$$

and

$$(\mathbf{i}\times\mathbf{j})\times\mathbf{j} = \mathbf{k}\times\mathbf{j} = -\mathbf{i}$$

so

$$\mathbf{i}\times(\mathbf{j}\times\mathbf{j}) \neq (\mathbf{i}\times\mathbf{j})\times\mathbf{j}$$

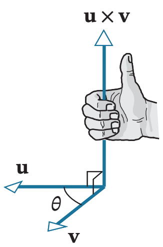


FIGURE 3.5.3

We know from Theorem 3.5.1 that $\mathbf{u}\times\mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$. If $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors, it can be shown that the direction of $\mathbf{u}\times\mathbf{v}$ can be determined using the following "right-hand rule" (Figure 3.5.3): Let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$, and suppose $\mathbf{u}$ is rotated through the angle θ until it coincides with $\mathbf{v}$. If the fingers of the right hand are cupped so that they point in the direction of rotation, then the thumb indicates (roughly) the direction of $\mathbf{u}\times\mathbf{v}$.

You may find it instructive to practice this rule with the products

$$\mathbf{i}\times\mathbf{j} = \mathbf{k}, \quad \mathbf{j}\times\mathbf{k} = \mathbf{i}, \quad \mathbf{k}\times\mathbf{i} = \mathbf{j}$$

Geometric Interpretation of Cross Product

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in 3-space, then the norm of $\mathbf{u}\times\mathbf{v}$ has a useful geometric interpretation. Lagrange's identity, given in Theorem 3.5.1, states that

$$\|\mathbf{u}\times\mathbf{v}\|^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - (\mathbf{u}\cdot\mathbf{v})^2 \tag{5}$$

If θ denotes the angle between $\mathbf{u}$ and $\mathbf{v}$, then $\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$, so (5) can be rewritten as

$$\begin{aligned} \|\mathbf{u}\times\mathbf{v}\|^2 &= \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - \|\mathbf{u}\|^2\|\mathbf{v}\|^2\cos^2\theta \\ &= \|\mathbf{u}\|^2\|\mathbf{v}\|^2(1-\cos^2\theta) \\ &= \|\mathbf{u}\|^2\|\mathbf{v}\|^2\sin^2\theta \end{aligned}$$

Since $0 \le \theta \le \pi$, it follows that $\sin\theta \ge 0$, so this can be rewritten as

$$\boxed{\|\mathbf{u}\times\mathbf{v}\| = \|\mathbf{u}\|\|\mathbf{v}\|\sin\theta} \tag{6}$$

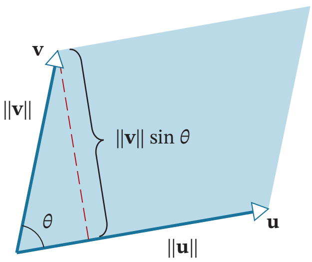


FIGURE 3.5.4

But $\|\mathbf{v}\|\sin\theta$ is the altitude of the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$ (Figure 3.5.4). Thus, from (6), the area A of this parallelogram is given by

$$A = (\text{base})(\text{altitude}) = \|\mathbf{u}\|\|\mathbf{v}\|\sin\theta = \|\mathbf{u}\times\mathbf{v}\|$$

This result is even correct if **u** and **v** are collinear, since the parallelogram determined by **u** and **v** has zero area and from (6) we have $\mathbf{u} \times \mathbf{v} = \mathbf{0}$ because $\theta = 0$ in this case. Thus we have the following theorem.

Theorem 3.5.3

Area of a Parallelogram

If **u** and **v** are vectors in 3-space, then $\|\mathbf{u} \times \mathbf{v}\|$ is equal to the area of the parallelogram determined by **u** and **v**.

EXAMPLE 4 | Area of a Triangle

Find the area of the triangle determined by the points $P_1(2,2,0)$, $P_2(-1,0,2)$, and $P_3(0,4,3)$.

Solution The area A of the triangle is $\frac{1}{2}$ the area of the parallelogram determined by the vectors $\overrightarrow{P_1P_2}$ and $\overrightarrow{P_1P_3}$ (**Figure 3.5.5**). Using the method discussed in Example 1 of Section 3.1, $\overrightarrow{P_1P_2} = (-3,-2,2)$ and $\overrightarrow{P_1P_3} = (-2,2,3)$. It follows that

$$\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3} = (-10, 5, -10)$$

(verify) and consequently that

$$A = \tfrac{1}{2}\|\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3}\| = \tfrac{1}{2}(15) = \tfrac{15}{2}$$

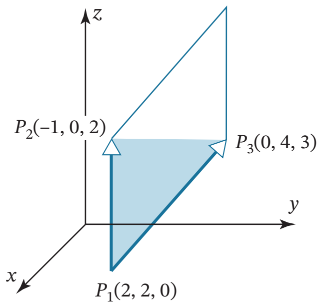


FIGURE 3.5.5

Definition 2

If **u**, **v**, and **w** are vectors in 3-space, then

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$$

is called the ***scalar triple product*** of **u**, **v**, and **w**.

The scalar triple product of $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$, and $\mathbf{w} = (w_1, w_2, w_3)$ can be calculated from the formula

$$\boxed{\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}} \tag{7}$$

This follows from Formula (4) since

$$\begin{aligned}
\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) &= \mathbf{u} \cdot \left(\begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix} \mathbf{k} \right) \\
&= \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} u_1 - \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} u_2 + \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix} u_3 \\
&= \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}
\end{aligned}$$

EXAMPLE 5 | Calculating a Scalar Triple Product

Calculate the scalar triple product $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$ of the vectors

$$\mathbf{u} = 3\mathbf{i} - 2\mathbf{j} - 5\mathbf{k}, \quad \mathbf{v} = \mathbf{i} + 4\mathbf{j} - 4\mathbf{k}, \quad \mathbf{w} = 3\mathbf{j} + 2\mathbf{k}$$

Solution From (7),

$$\begin{aligned}
\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w}) &= \begin{vmatrix} 3 & -2 & -5 \\ 1 & 4 & -4 \\ 0 & 3 & 2 \end{vmatrix} \\
&= 3\begin{vmatrix} 4 & -4 \\ 3 & 2 \end{vmatrix} - (-2)\begin{vmatrix} 1 & -4 \\ 0 & 2 \end{vmatrix} + (-5)\begin{vmatrix} 1 & 4 \\ 0 & 3 \end{vmatrix} \\
&= 60 + 4 - 15 = 49
\end{aligned}$$

Remark The symbol $(\mathbf{u}\cdot\mathbf{v})\times\mathbf{w}$ makes no sense because we cannot form the cross product of a scalar and a vector. Thus, no ambiguity arises if we write $\mathbf{u}\cdot\mathbf{v}\times\mathbf{w}$ rather than $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$. However, for clarity we will usually keep the parentheses.

It follows from (7) that

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

since the 3×3 determinants that represent these products can be obtained from one another by *two* row interchanges. (Verify.) These relationships can be remembered by moving the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ clockwise around the vertices of the triangle in Figure 3.5.6.

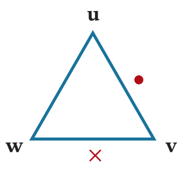


FIGURE 3.5.6

Geometric Interpretation of Determinants

The next theorem provides a useful geometric interpretation of 2×2 and 3×3 determinants.

Theorem 3.5.4

(*a*) The absolute value of the determinant

$$\det\begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix}$$

is equal to the area of the parallelogram in 2-space determined by the vectors $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$. (See Figure 3.5.7*a*.)

(*b*) The absolute value of the determinant

$$\det\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix}$$

is equal to the volume of the parallelepiped in 3-space determined by the vectors $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$, and $\mathbf{w} = (w_1, w_2, w_3)$. (See Figure 3.5.7*b*.)

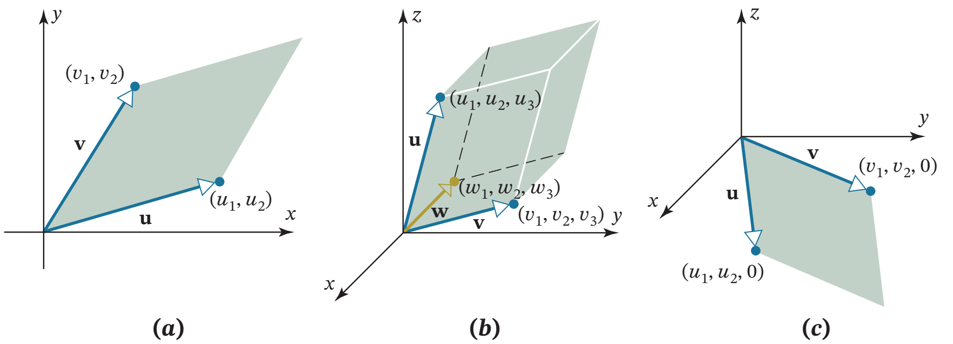


FIGURE 3.5.7

Proof (a) The key to the proof is to use Theorem 3.5.3. However, that theorem applies to vectors in 3-space, whereas $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ are vectors in 2-space. To circumvent this "dimension problem," we will view $\mathbf{u}$ and $\mathbf{v}$ as vectors in the xy-plane of an xyz-coordinate system (**Figure 3.5.7c**), in which case these vectors are expressed as $\mathbf{u} = (u_1, u_2, 0)$ and $\mathbf{v} = (v_1, v_2, 0)$. Thus

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & 0 \\ v_1 & v_2 & 0 \end{vmatrix} = \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \mathbf{k} = \det \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \mathbf{k}$$

It now follows from Theorem 3.5.3 and the fact that $\|\mathbf{k}\| = 1$ that the area A of the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$ is

$$A = \|\mathbf{u} \times \mathbf{v}\| = \left\| \det \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \mathbf{k} \right\| = \left| \det \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \right| \|\mathbf{k}\| = \left| \det \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \right|$$

which completes the proof.

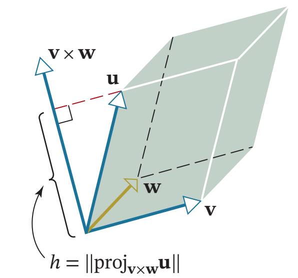


FIGURE 3.5.8

Proof (b) As shown in **Figure 3.5.8**, take the base of the parallelepiped determined by $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ to be the parallelogram determined by $\mathbf{v}$ and $\mathbf{w}$. It follows from Theorem 3.5.3 that the area of the base is $\|\mathbf{v} \times \mathbf{w}\|$ and, as illustrated in Figure 3.5.8, the height h of the parallelepiped is the length of the orthogonal projection of $\mathbf{u}$ on $\mathbf{v} \times \mathbf{w}$. Therefore, by Formula (12) of Section 3.3,

$$h = \|\text{proj}_{\mathbf{v}\times\mathbf{w}}\mathbf{u}\| = \frac{|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|}{\|\mathbf{v} \times \mathbf{w}\|}$$

It follows that the volume V of the parallelepiped is

$$V = (\text{area of base}) \cdot \text{height} = \|\mathbf{v} \times \mathbf{w}\| \frac{|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|}{\|\mathbf{v} \times \mathbf{w}\|} = |\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|$$

so from (7),

$$V = \left| \det \begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix} \right| \tag{8}$$

which completes the proof. ■

Remark If V denotes the volume of the parallelepiped determined by vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$, then it follows from Formulas (7) and (8) that

$$\boxed{V = \begin{bmatrix} \text{volume of parallelepiped} \\ \text{determined by } \mathbf{u}, \mathbf{v}, \text{and } \mathbf{w} \end{bmatrix} = |\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|} \tag{9}$$

From this result and the discussion immediately following Definition 3 of Section 3.2, we can conclude that

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

where the + or − results depending on whether $\mathbf{u}$ makes an acute or an obtuse angle with $\mathbf{v} \times \mathbf{w}$.

Formula (9) leads to a useful test for ascertaining whether three given vectors lie in the same plane. Since three vectors not in the same plane determine a parallelepiped of positive volume, it follows from (9) that $|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})| = 0$ if and only if the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ lie in the same plane. Thus we have the following result.

Theorem 3.5.5

If the vectors $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$, and $\mathbf{w} = (w_1, w_2, w_3)$ have the same initial point, then they lie in the same plane if and only if

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = 0$$

Exercise Set 3.5

In Exercises 1–2, let $\mathbf{u} = (3, 2, -1)$, $\mathbf{v} = (0, 2, -3)$, *and* $\mathbf{w} = (2, 6, 7)$. *Compute the indicated vectors.*

1. **a.** $\mathbf{v} \times \mathbf{w}$ **b.** $\mathbf{w} \times \mathbf{v}$ **c.** $(\mathbf{u} + \mathbf{v}) \times \mathbf{w}$
d. $\mathbf{v} \cdot (\mathbf{v} \times \mathbf{w})$ **e.** $\mathbf{v} \times \mathbf{v}$ **f.** $(\mathbf{u} - 3\mathbf{w}) \times (\mathbf{u} - 3\mathbf{w})$

2. **a.** $\mathbf{u} \times \mathbf{v}$ **b.** $-(\mathbf{u} \times \mathbf{v})$ **c.** $\mathbf{u} \times (\mathbf{v} + \mathbf{w})$
d. $\mathbf{w} \cdot (\mathbf{w} \times \mathbf{v})$ **e.** $\mathbf{w} \times \mathbf{w}$ **f.** $(7\mathbf{v} - 3\mathbf{u}) \times (7\mathbf{v} - 3\mathbf{u})$

In Exercises 3–4, let $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *be the vectors in Exercises 1–2. Use Lagrange's identity to rewrite the expression using only dot products and scalar multiplications, and then confirm your result by evaluating both sides of the identity.*

3. $\|\mathbf{u} \times \mathbf{w}\|^2$ **4.** $\|\mathbf{v} \times \mathbf{u}\|^2$

In Exercises 5–6, let $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *be the vectors in Exercises 1–2. Compute the vector triple product directly, and check your result by using parts (d) and (e) of Theorem 3.5.1.*

5. $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ **6.** $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$

In Exercises 7–8, use the cross product to find a vector that is orthogonal to both $\mathbf{u}$ *and* $\mathbf{v}$.

7. $\mathbf{u} = (-6, 4, 2)$, $\mathbf{v} = (3, 1, 5)$

8. $\mathbf{u} = (1, 1, -2)$, $\mathbf{v} = (2, -1, 2)$

In Exercises 9–10, find the area of the parallelogram determined by the given vectors $\mathbf{u}$ *and* $\mathbf{v}$.

9. $\mathbf{u} = (1, -1, 2)$, $\mathbf{v} = (0, 3, 1)$

10. $\mathbf{u} = (3, -1, 4)$, $\mathbf{v} = (6, -2, 8)$

In Exercises 11–12, find the area of the parallelogram with the given vertices.

11. $P_1(1, 2)$, $P_2(4, 4)$, $P_3(7, 5)$, $P_4(4, 3)$

12. $P_1(3, 2)$, $P_2(5, 4)$, $P_3(9, 4)$, $P_4(7, 2)$

In Exercises 13–14, find the area of the triangle with the given vertices.

13. $A(2, 0)$, $B(3, 4)$, $C(-1, 2)$

14. $A(1, 1)$, $B(2, 2)$, $C(3, -3)$

In Exercises 15–16, find the area of the triangle in 3-space that has the given vertices.

15. $P_1(2, 6, -1)$, $P_2(1, 1, 1)$, $P_3(4, 6, 2)$

16. $P(1, -1, 2)$, $Q(0, 3, 4)$, $R(6, 1, 8)$

In Exercises 17–18, find the volume of the parallelepiped with sides $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$.

17. $\mathbf{u} = (2, -6, 2)$, $\mathbf{v} = (0, 4, -2)$, $\mathbf{w} = (2, 2, -4)$

18. $\mathbf{u} = (3, 1, 2)$, $\mathbf{v} = (4, 5, 1)$, $\mathbf{w} = (1, 2, 4)$

In Exercises 19–20, determine whether $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *lie in the same plane when positioned so that their initial points coincide.*

19. $\mathbf{u} = (-1, -2, 1)$, $\mathbf{v} = (3, 0, -2)$, $\mathbf{w} = (5, -4, 0)$

20. $\mathbf{u} = (5, -2, 1)$, $\mathbf{v} = (4, -1, 1)$, $\mathbf{w} = (1, -1, 0)$

In Exercises 21–24, compute the scalar triple product $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$.

21. $\mathbf{u} = (-2, 0, 6)$, $\mathbf{v} = (1, -3, 1)$, $\mathbf{w} = (-5, -1, 1)$

22. $\mathbf{u} = (-1, 2, 4)$, $\mathbf{v} = (3, 4, -2)$, $\mathbf{w} = (-1, 2, 5)$

23. $\mathbf{u} = (a, 0, 0)$, $\mathbf{v} = (0, b, 0)$, $\mathbf{w} = (0, 0, c)$

24. $\mathbf{u} = \mathbf{i}$, $\mathbf{v} = \mathbf{j}$, $\mathbf{w} = \mathbf{k}$

In Exercises 25–26, suppose that $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = 3$. *Find*

25. **a.** $\mathbf{u} \cdot (\mathbf{w} \times \mathbf{v})$ **b.** $(\mathbf{v} \times \mathbf{w}) \cdot \mathbf{u}$ **c.** $\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v})$

26. **a.** $\mathbf{v} \cdot (\mathbf{u} \times \mathbf{w})$ **b.** $(\mathbf{u} \times \mathbf{w}) \cdot \mathbf{v}$ **c.** $\mathbf{v} \cdot (\mathbf{w} \times \mathbf{w})$

27. **a.** Find the area of the triangle having vertices $A(1, 0, 1)$, $B(0, 2, 3)$, and $C(2, 1, 0)$.

b. Use the result of part (a) to find the length of the altitude from vertex C to side AB.

28. Use the cross product to find the sine of the angle between the vectors $\mathbf{u} = (2, 3, -6)$ and $\mathbf{v} = (2, 3, 6)$.

29. Simplify $(\mathbf{u} + \mathbf{v}) \times (\mathbf{u} - \mathbf{v})$.

30. Let $\mathbf{a} = (a_1, a_2, a_3)$, $\mathbf{b} = (b_1, b_2, b_3)$, $\mathbf{c} = (c_1, c_2, c_3)$, and $\mathbf{d} = (d_1, d_2, d_3)$. Show that

$$(\mathbf{a} + \mathbf{d}) \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) + \mathbf{d} \cdot (\mathbf{b} \times \mathbf{c})$$

Exercises 31–32 *You know from your own experience that the tendency for a force to cause a rotation about an axis depends on the amount of force applied and its distance from the axis of rotation. For example, it is easier to close a door by pushing on its outer edge than close to its hinges. Moreover, the harder you push, the faster the door will close. In physics, the tendency for a force vector* $\mathbf{F}$ *to cause rotational motion is a vector called* **torque** *(denoted by* $\boldsymbol{\tau}$*). It is defined as*

$$\boldsymbol{\tau} = \mathbf{F} \times \mathbf{d}$$

where $\mathbf{d}$ *is the vector from the axis of rotation to the point at which the force is applied. It follows from Formula (6) that*

$$\|\boldsymbol{\tau}\| = \|\mathbf{F} \times \mathbf{d}\| = \|\mathbf{F}\|\|\mathbf{d}\| \sin\theta$$

where θ *is the angle between the vectors* $\mathbf{F}$ *and* $\mathbf{d}$*. This is called the* **scalar moment** *of* $\mathbf{F}$ *about the axis of rotation and is typically measured in units of Newton meters (Nm) or foot pounds (ft–lb).*

31. The accompanying figure shows a force $\mathbf{F}$ of 1000 N applied to the corner of a box.

a. Find the scalar moment of $\mathbf{F}$ about the point P.

b. Find the direction angles of the vector moment of $\mathbf{F}$ about the point P to the nearest degree. [See directions for Exercises 21–25 of Section 3.2.]

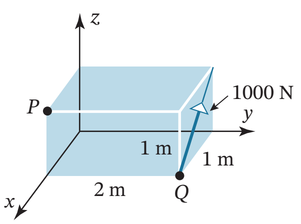


FIGURE Ex-31

32. As shown in the accompanying figure, a force of 200 N is applied at an angle of 18° to a point near the end of a monkey wrench. Find the scalar moment of the force about the center of the bolt. [*Note:* Treat the wrench as two-dimensional.]

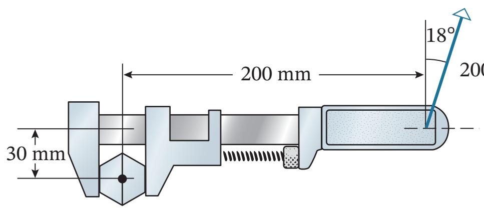


FIGURE Ex-32

Working with Proofs

33. Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be nonzero vectors in 3-space with the same initial point, but such that no two of them are collinear. Prove that

a. $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ lies in the plane determined by $\mathbf{v}$ and $\mathbf{w}$.

b. $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$ lies in the plane determined by $\mathbf{u}$ and $\mathbf{v}$.

34. Prove the following identities.

a. $(\mathbf{u} + k\mathbf{v}) \times \mathbf{v} = \mathbf{u} \times \mathbf{v}$

b. $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{z}) = -(\mathbf{u} \times \mathbf{z}) \cdot \mathbf{v}$

35. Prove: If $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ lie in the same plane, then $(\mathbf{a} \times \mathbf{b}) \times (\mathbf{c} \times \mathbf{d}) = \mathbf{0}$.

36. Prove: If θ is the angle between $\mathbf{u}$ and $\mathbf{v}$ and $\mathbf{u} \cdot \mathbf{v} \neq 0$, then $\tan\theta = \|\mathbf{u} \times \mathbf{v}\|/(\mathbf{u} \cdot \mathbf{v})$.

37. Prove that if $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in R^3, no two of which are collinear, then $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ lies in the plane determined by $\mathbf{v}$ and $\mathbf{w}$.

38. It is a theorem of solid geometry that the volume of a tetrahedron is $\frac{1}{3}$(area of base) $\cdot$ (height). Use this result to prove that the volume of a tetrahedron whose sides are the vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ is $\frac{1}{6}|\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})|$ (see accompanying figure).

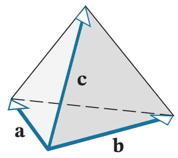


FIGURE Ex-38

39. Use the result of Exercise 38 to find the volume of the tetrahedron with vertices P, Q, R, S.

a. $P(-1, 2, 0)$, $Q(2, 1, -3)$, $R(1, 1, 1)$, $S(3, -2, 3)$

b. $P(0, 0, 0)$, $Q(1, 2, -1)$, $R(3, 4, 0)$, $S(-1, -3, 4)$

40. Prove part (*d*) of Theorem 3.5.1. [*Hint:* First prove the result in the case where $\mathbf{w} = \mathbf{i} = (1, 0, 0)$, then when $\mathbf{w} = \mathbf{j} = (0, 1, 0)$, and then when $\mathbf{w} = \mathbf{k} = (0, 0, 1)$. Finally, prove it for an arbitrary vector $\mathbf{w} = (w_1, w_2, w_3)$ by writing $\mathbf{w} = w_1\mathbf{i} + w_2\mathbf{j} + w_3\mathbf{k}$.]

41. Prove part (*e*) of Theorem 3.5.1. [*Hint:* Apply part (*a*) of Theorem 3.5.2 to the result in part (*d*) of Theorem 3.5.1.]

42. Prove:

a. Prove (*b*) of Theorem 3.5.2.

b. Prove (*c*) of Theorem 3.5.2.

c. Prove (*d*) of Theorem 3.5.2.

d. Prove (*e*) of Theorem 3.5.2.

e. Prove (*f*) of Theorem 3.5.2.

True-False Exercises

TF. In parts **(a)–(f)** determine whether the statement is true or false, and justify your answer.

a. The cross product of two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ is a nonzero vector if and only if $\mathbf{u}$ and $\mathbf{v}$ are not parallel.

b. A normal vector to a plane can be obtained by taking the cross product of two nonzero and noncollinear vectors lying in the plane.

c. The scalar triple product of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ determines a vector whose length is equal to the volume of the parallelepiped determined by $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$.

d. If $\mathbf{u}$ and $\mathbf{v}$ are vectors in 3-space, then $\|\mathbf{v}\times\mathbf{u}\|$ is equal to the area of the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$.

e. For all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in 3-space, the vectors

$$(\mathbf{u}\times\mathbf{v})\times\mathbf{w} \quad \text{and} \quad \mathbf{u}\times(\mathbf{v}\times\mathbf{w})$$

are the same.

f. If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in R^3, where $\mathbf{u}$ is nonzero and $\mathbf{u}\times\mathbf{v}=\mathbf{u}\times\mathbf{w}$, then $\mathbf{v}=\mathbf{w}$.

Working with Technology

T1. As stated in Exercise 23, the distance d in 3-space from a point P to the line L through points A and B is given by the formula

$$d = \frac{\|\overrightarrow{AP}\times\overrightarrow{AB}\|}{\|\overrightarrow{AB}\|}$$

Find the distance between the point $P(1,3,1)$ and the line through the points $A(2,-3,4)$ and $B(4,7,-2)$.

Chapter 3 Supplementary Exercises

1. Let $\mathbf{u}=(-2,0,4)$, $\mathbf{v}=(3,-1,6)$, and $\mathbf{w}=(2,-5,-5)$. Compute

a. $3\mathbf{v}-2\mathbf{u}$ **b.** $\|\mathbf{u}+\mathbf{v}+\mathbf{w}\|$

c. the distance between $-3\mathbf{u}$ and $\mathbf{v}+5\mathbf{w}$

d. $\text{proj}_{\mathbf{w}}\mathbf{u}$ **e.** $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$

f. $(-5\mathbf{v}+\mathbf{w})\times((\mathbf{u}\cdot\mathbf{v})\mathbf{w})$

2. Repeat Exercise 1 for the vectors

$$\mathbf{u}=3\mathbf{i}-5\mathbf{j}+\mathbf{k}, \mathbf{v}=-2\mathbf{i}+2\mathbf{k},$$

and

$$\mathbf{w}=-\mathbf{j}+4\mathbf{k}$$

3. Repeat parts (a)–(d) of Exercise 1 using the vectors $\mathbf{u}=(-2,6,2,1)$, $\mathbf{v}=(-3,0,8,0)$, and $\mathbf{w}=(9,1,-6,-6)$.

4. **a.** The set of all vectors in R^2 that are orthogonal to a nonzero vector is what kind of geometric object?

b. The set of all vectors in R^3 that are orthogonal to a nonzero vector is what kind of geometric object?

c. The set of all vectors in R^2 that are orthogonal to two noncollinear vectors is what kind of geometric object?

d. The set of all vectors in R^3 that are orthogonal to two noncollinear vectors is what kind of geometric object?

5. Let A, B, and C be three distinct noncollinear points in 3-space. Describe the set of all points P that satisfy the vector equation $\overrightarrow{AP}\cdot(\overrightarrow{AB}\times\overrightarrow{AC})=0$.

6. Let A, B, C, and D be four distinct noncollinear points in 3-space. If $\overrightarrow{AB}\times\overrightarrow{CD}\neq\mathbf{0}$ and $\overrightarrow{AC}\cdot(\overrightarrow{AB}\times\overrightarrow{CD})=0$, explain why the line through A and B must intersect the line through C and D.

7. Consider the points $P(3,-1,4)$, $Q(6,0,2)$, and $R(5,1,1)$. Find the point S in R^3 whose first component is -1 and such that $\overrightarrow{PQ}$ is parallel to $\overrightarrow{RS}$.

8. Consider the points $P(-3,1,0,6)$, $Q(0,5,1,-2)$, and $R(-4,1,4,0)$. Find the point S in R^4 whose third component is 6 and such that $\overrightarrow{PQ}$ is parallel to $\overrightarrow{RS}$.

9. Using the points in Exercise 7, find the cosine of the angle between the vectors $\overrightarrow{PQ}$ and $\overrightarrow{PR}$.

10. Using the points in Exercise 8, find the cosine of the angle between the vectors $\overrightarrow{PQ}$ and $\overrightarrow{PR}$.

11. Find the distance between the point $P(-3,1,3)$ and the plane $5x+z=3y-4$.

12. Show that the planes

$$3x-y+6z=7$$

and

$$-6x+2y-12z=1$$

are parallel, and find the distance between them.

In Exercises 13–18, find vector and parametric equations for the line or plane in question.

13. The plane in R^3 that contains the points $P(-2,1,3)$, $Q(-1,-1,1)$, and $R(3,0,-2)$.

14. The line in R^3 that contains the point $P(-1,6,0)$ and is orthogonal to the plane $4x-z=5$.

15. The line in R^2 that is parallel to the vector $\mathbf{v}=(8,-1)$ and contains the point $P(0,-3)$.

16. The plane in R^3 that contains the point $P(-2,1,0)$ and is parallel to the plane $-8x+6y-z=4$.

17. The line in R^2 with equation $y=3x-5$.

18. The plane in R^3 with equation $2x-6y+3z=5$.

In Exercises 19–21, find a point-normal equation for the given plane.

19. The plane that is represented by the vector equation

$$(x,y,z)=(-1,5,6)+t_1(0,-1,3)+t_2(2,-1,0)$$

20. The plane that contains the point $P(-5, 1, 0)$ and is orthogonal to the line with parametric equations $x = 3 - 5t$, $y = 2t$, and $z = 7$.

21. The plane that passes through the points $P(9, 0, 4)$, $Q(-1, 4, 3)$, and $R(0, 6, -2)$.

22. Suppose that $V = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $W = \{\mathbf{w}_1, \mathbf{w}_2\}$ are two sets of vectors such that each vector in V is orthogonal to each vector in W. Prove that if a_1, a_2, a_3, b_1, b_2 are any scalars, then the vectors $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3$ and $\mathbf{w} = b_1\mathbf{w}_1 + b_2\mathbf{w}_2$ are orthogonal.

23. Show that in 3-space the distance d from a point P to the line L through points A and B can be expressed as

$$d = \frac{\|\overrightarrow{AP} \times \overrightarrow{AB}\|}{\|\overrightarrow{AB}\|}$$

24. Prove that $\|\mathbf{u} + \mathbf{v}\| = \|\mathbf{u}\| + \|\mathbf{v}\|$ if and only if one of the vectors is a scalar multiple of the other.

25. The equation $Ax + By = 0$ represents a line through the origin in R^2 if A and B are not both zero. What does this equation represent in R^3 if you think of it as $Ax + By + 0z = 0$? Explain.

CHAPTER 4

General Vector Spaces

CHAPTER CONTENTS

Introduction

Recall that we began our study of vectors by viewing them as directed line segments (arrows). We then extended this idea by introducing rectangular coordinate systems, and that enabled us to view vectors as ordered pairs and ordered triples of real numbers. As we developed properties of these vectors we noticed patterns in various formulas that enabled us to extend the notion of a vector to an n-tuple of real numbers. Although n-tuples took us outside the realm of our "visual experience," it gave us a valuable tool for understanding and studying systems of linear equations. In this chapter we will extend the concept of a vector yet again by using the most important algebraic properties of vectors in R^n as axioms. These axioms, if satisfied by a set of objects, will enable us to think of those objects as vectors.

4.1 Real Vector Spaces

In this section we will extend the concept of a vector by using the basic properties of vectors in R^n as axioms, which if satisfied by a set of objects will guarantee that those objects behave like familiar vectors.

Vector Space Axioms

The following definition consists of ten axioms, eight of which are properties of vectors in R^n that were stated in Theorem 3.1.1. It is important to keep in mind that one does not *prove* axioms; rather, they are assumptions that serve as the starting point for proving theorems.

Definition 1

Let V be an arbitrary nonempty set of objects for which two operations are defined: addition and multiplication by numbers called ***scalars***. By ***addition*** we mean a rule for associating with each pair of objects $\mathbf{u}$ and $\mathbf{v}$ in V an object $\mathbf{u}+\mathbf{v}$, called the ***sum*** of $\mathbf{u}$ and $\mathbf{v}$; by ***scalar multiplication*** we mean a rule for associating with each scalar k and each object $\mathbf{u}$ in V an object $k\mathbf{u}$, called the ***scalar multiple*** of $\mathbf{u}$ by k. If the following axioms are satisfied by all objects $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in V and all scalars k and m, then we call V a ***vector space*** and we call the objects in V ***vectors***.

1. If $\mathbf{u}$ and $\mathbf{v}$ are objects in V, then $\mathbf{u}+\mathbf{v}$ is in V.
2. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$
3. $\mathbf{u}+(\mathbf{v}+\mathbf{w})=(\mathbf{u}+\mathbf{v})+\mathbf{w}$
4. There exists an object in V, called the ***zero vector***, that is denoted by $\mathbf{0}$ and has the property that $\mathbf{0}+\mathbf{u}=\mathbf{u}+\mathbf{0}=\mathbf{u}$ for all $\mathbf{u}$ in V.
5. For each $\mathbf{u}$ in V, there is an object $-\mathbf{u}$ in V, called a ***negative*** of $\mathbf{u}$, such that $\mathbf{u}+(-\mathbf{u})=(-\mathbf{u})+\mathbf{u}=\mathbf{0}$.
6. If k is any scalar and $\mathbf{u}$ is any object in V, then $k\mathbf{u}$ is in V.
7. $k(\mathbf{u}+\mathbf{v})=k\mathbf{u}+k\mathbf{v}$
8. $(k+m)\mathbf{u}=k\mathbf{u}+m\mathbf{u}$
9. $k(m\mathbf{u})=(km)(\mathbf{u})$
10. $1\mathbf{u}=\mathbf{u}$

In this text scalars will be either real numbers or complex numbers. Vector spaces with real scalars will be called ***real vector spaces*** and those with complex scalars will be called ***complex vector spaces***. For now we will consider only real vector spaces.

Observe that the definition of a vector space does not specify the nature of the vectors or the operations. Any kind of object can be a vector, and the operations of addition and scalar multiplication need not have any relationship to those on R^n. The only requirement is that the ten vector space axioms be satisfied. In the examples that follow we will use four basic steps to show that a set with two operations is a vector space.

Steps to Show That a Set with Two Operations Is a Vector Space

Step 1. Identify the set V of objects that will become vectors.

Step 2. Identify the addition and scalar multiplication operations on V.

Step 3. Verify Axioms 1 and 6; that is, adding two vectors in V produces a vector in V, and multiplying a vector in V by a scalar also produces a vector in V.

Axiom 1 is called ***closure under addition***, and Axiom 6 is called ***closure under scalar multiplication***.

Step 4. Confirm that Axioms 2, 3, 4, 5, 7, 8, 9, and 10 hold.

Our first example is the simplest of all vector spaces in that it contains only one object. Since Axiom 4 requires that every vector space contain a zero vector, the object will have to be that vector.

EXAMPLE 1 | The Zero Vector Space

Let V consist of a single object, which we denote by $\mathbf{0}$, and define

$$\mathbf{0} + \mathbf{0} = \mathbf{0} \quad \text{and} \quad k\mathbf{0} = \mathbf{0}$$

for all scalars k. It is easy to check that all the vector space axioms are satisfied. We call this the ***zero vector space***.

Our second example is one of the most important of all vector spaces—the familiar space R^n. It should not be surprising that the operations on R^n satisfy the vector space axioms because those axioms were based on known properties of operations on R^n.

EXAMPLE 2 | R^n Is a Vector Space

Let $V = R^n$, and define the vector space operations on V to be the usual operations of addition and scalar multiplication of n-tuples; that is,

$$\mathbf{u} + \mathbf{v} = (u_1, u_2, \ldots, u_n) + (v_1, v_2, \ldots, v_n) = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n)$$
$$k\mathbf{u} = (ku_1, ku_2, \ldots, ku_n)$$

The set $V = R^n$ is closed under addition and scalar multiplication because the foregoing operations produce n-tuples as their end result, and these operations satisfy Axioms 2, 3, 4, 5, 7, 8, 9, and 10 by virtue of Theorem 3.1.1.

Historical Note

Hermann Günther Grassmann (1809–1877)

The notion of an "abstract vector space" evolved over many years and had many contributors. The idea crystallized with the work of the German mathematician H. G. Grassmann, who published a paper in 1862 in which he considered abstract systems of unspecified elements on which he defined formal operations of addition and scalar multiplication. Grassmann's work was controversial, and others, including Augustin Cauchy (p. 136), laid reasonable claim to the idea.

Our next example is a generalization of R^n in which we allow vectors to have infinitely many components.

EXAMPLE 3 | The Vector Space of Infinite Sequences of Real Numbers

Let V consist of objects of the form

$$\mathbf{u} = (u_1, u_2, \ldots, u_n, \ldots)$$

in which $u_1, u_2, \ldots, u_n, \ldots$ is an infinite sequence of real numbers. We define two infinite sequences to be *equal* if their corresponding components are equal, and we define addition and scalar multiplication componentwise by

$$\begin{aligned}\mathbf{u} + \mathbf{v} &= (u_1, u_2, \ldots, u_n, \ldots) + (v_1, v_2, \ldots, v_n, \ldots)\\ &= (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n, \ldots)\\ k\mathbf{u} &= (ku_1, ku_2, \ldots, ku_n, \ldots)\end{aligned}$$

In the exercises we ask you to confirm that V with these operations is a vector space. We will denote this vector space by the symbol R^∞.

Vector spaces of the type in Example 3 arise when a transmitted signal of indefinite duration is digitized by sampling its values at discrete time intervals (**Figure 4.1.1**).

In the next example our vectors will be matrices. This may be a little confusing at first because matrices are composed of rows and columns, which are themselves vectors (row vectors and column vectors). However, from the vector space viewpoint we are not concerned with the individual rows and columns but rather with the properties of the matrix operations as they relate to the matrix as a whole.

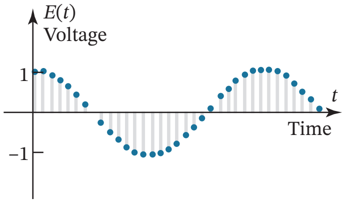


FIGURE 4.1.1 This transmitted signal continues indefinitely.

EXAMPLE 4 | The Vector Space of 2×2 Matrices

Let V be the set of 2×2 matrices with real entries, and take the vector space operations on V to be the usual operations of matrix addition and scalar multiplication; that is,

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} + \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} = \begin{bmatrix} u_{11} + v_{11} & u_{12} + v_{12} \\ u_{21} + v_{21} & u_{22} + v_{22} \end{bmatrix} \tag{1}$$

$$k\mathbf{u} = k\begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} = \begin{bmatrix} ku_{11} & ku_{12} \\ ku_{21} & ku_{22} \end{bmatrix}$$

Note that Equation (1) involves *three* different addition operations: the addition operation on vectors, the addition operation on matrices, and the addition operation on real numbers.

The set V is closed under addition and scalar multiplication because the foregoing operations produce 2×2 matrices as the end result. Thus, it remains to confirm that Axioms 2, 3, 4, 5, 7, 8, 9, and 10 hold. Some of these are standard properties of matrix operations. For example, Axiom 2 follows from Theorem 1.4.1(a) since

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} + \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} + \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} = \mathbf{v} + \mathbf{u}$$

Similarly, Axioms 3, 7, 8, and 9 follow from parts (b), (h), (j), and (e), respectively, of that theorem (verify). This leaves Axioms 4, 5, and 10 that remain to be verified.

To confirm that Axiom 4 is satisfied, we must find a 2×2 matrix $\mathbf{0}$ in V for which $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u}$ for all 2×2 matrices in V. We can do this by taking

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

With this definition,

$$\mathbf{0} + \mathbf{u} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} = \mathbf{u}$$

and similarly $\mathbf{u} + \mathbf{0} = \mathbf{u}$. To verify that Axiom 5 holds we must show that each object $\mathbf{u}$ in V has a negative $-\mathbf{u}$ in V such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$ and $(-\mathbf{u}) + \mathbf{u} = \mathbf{0}$. This can be done by defining the negative of $\mathbf{u}$ to be

$$-\mathbf{u} = \begin{bmatrix} -u_{11} & -u_{12} \\ -u_{21} & -u_{22} \end{bmatrix}$$

With this definition,

$$\mathbf{u} + (-\mathbf{u}) = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} + \begin{bmatrix} -u_{11} & -u_{12} \\ -u_{21} & -u_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \mathbf{0}$$

and similarly $(-\mathbf{u}) + \mathbf{u} = \mathbf{0}$. Finally, Axiom 10 holds because

$$1\mathbf{u} = 1\begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} = \mathbf{u}$$

EXAMPLE 5 | The Vector Space of $m \times n$ Matrices

Example 4 is a special case of a more general class of vector spaces. You should have no trouble adapting the argument used in that example to show that the set V of all $m \times n$ matrices with the usual matrix operations of addition and scalar multiplication is a vector space. We will denote this vector space by the symbol M_{mn}. Thus, for example, the vector space in Example 4 is denoted as M_{22}.

EXAMPLE 6 | The Vector Space of Real-Valued Functions

Let V be the set of real-valued functions that are defined at each x in the interval $(-\infty, \infty)$. If $\mathbf{f} = f(x)$ and $\mathbf{g} = g(x)$ are two functions in V and if k is any scalar, then define the operations of addition and scalar multiplication by

$$(\mathbf{f} + \mathbf{g})(x) = f(x) + g(x) \tag{2}$$

$$(k\mathbf{f})(x) = kf(x) \tag{3}$$

One way to think about these operations is to view the numbers $f(x)$ and $g(x)$ as "components" of $\mathbf{f}$ and $\mathbf{g}$ at the point x, in which case Equations (2) and (3) state that two functions are added by adding corresponding components, and a function is multiplied by a scalar by multiplying each component by that scalar—exactly as in R^n and R^∞. This idea is illustrated in parts (a) and (b) of **Figure 4.1.2**. The set V with these operations is denoted by the symbol $F(-\infty, \infty)$. We can prove that this is a vector space as follows:

In Example 6 the functions are defined on the entire interval $(-\infty, \infty)$. However, the arguments used in that example apply as well on all subintervals of $(-\infty, \infty)$, such as a closed interval $[a, b]$ or an open interval (a, b). We will denote the vector spaces of functions on these intervals by $F[a, b]$ and $F(a, b)$, respectively.

Axioms 1 and 6: These closure axioms require that if we add two functions that are defined at each x in the interval $(-\infty, \infty)$, then sums and scalar multiples of those functions must also be defined at each x in the interval $(-\infty, \infty)$. This follows from Formulas (2) and (3).

Axiom 4: This axiom requires that there exists a function $\mathbf{0}$ in $F(-\infty, \infty)$, which when added to any other function $\mathbf{f}$ in $F(-\infty, \infty)$ produces $\mathbf{f}$ back again as the result. The function whose value at every point x in the interval $(-\infty, \infty)$ is zero has this property. Geometrically, the graph of the function $\mathbf{0}$ is the line that coincides with the x-axis.

Axiom 5: This axiom requires that for each function $\mathbf{f}$ in $F(-\infty, \infty)$ there exists a function $-\mathbf{f}$ in $F(-\infty, \infty)$, which when added to $\mathbf{f}$ produces the function $\mathbf{0}$. The function defined by $-\mathbf{f}(x) = -f(x)$ has this property. The graph of $-\mathbf{f}$ can be obtained by reflecting the graph of $\mathbf{f}$ about the x-axis (**Figure 4.1.2c**).

Axioms 2, 3, 7, 8, 9, 10: The validity of each of these axioms follows from properties of real numbers. For example, if $\mathbf{f}$ and $\mathbf{g}$ are functions in $F(-\infty, \infty)$, then Axiom 2 requires that $\mathbf{f} + \mathbf{g} = \mathbf{g} + \mathbf{f}$. This follows from the computation

$$(\mathbf{f} + \mathbf{g})(x) = f(x) + g(x) = g(x) + f(x) = (\mathbf{g} + \mathbf{f})(x)$$

in which the first and last equalities follow from (2), and the middle equality is a property of real numbers. We will leave the proofs of the remaining parts as exercises.

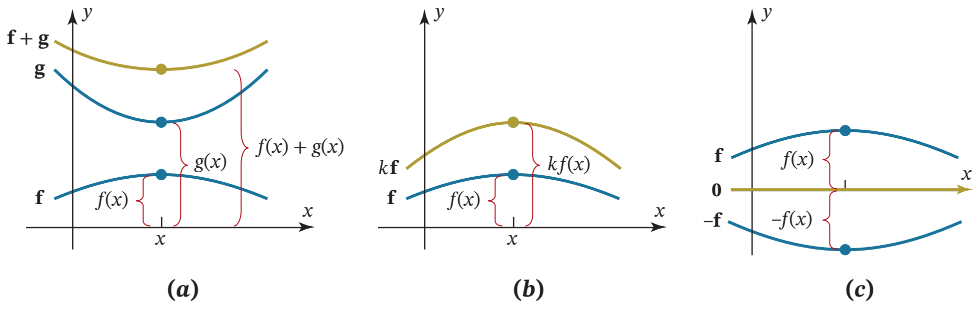


FIGURE 4.1.2

It is important to recognize that you cannot impose any two operations on any set V and expect the vector space axioms to hold. For example, if V is the set of n-tuples with *positive* components, and if the standard operations from R^n are used, then V is not closed under scalar multiplication because if $\mathbf{u}$ is a nonzero n-tuple in V, then $(-1)\mathbf{u}$ has at least one negative component and hence is not in V. The following is a less obvious example in which only one of the ten vector space axioms fails to hold.

EXAMPLE 7 | A Set That Is Not a Vector Space

Let $V = R^2$ and define addition and scalar multiplication operations as follows: If $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$, then define

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2)$$

and if k is any real number, then define

$$k\mathbf{u} = (ku_1, 0)$$

For example, if $\mathbf{u} = (2, 4)$, $\mathbf{v} = (-3, 5)$, and $k = 7$, then

$$\mathbf{u} + \mathbf{v} = (-1, 9)$$
$$k\mathbf{u} = 7\mathbf{u} = (14, 0)$$

The addition operation is the standard one from R^2, but the scalar multiplication is not. In the exercises we will ask you to show that the first nine vector space axioms are satisfied, but Axiom 10 fails to hold for certain vectors. For example, if $\mathbf{u} = (u_1, u_2)$ is such that $u_2 \neq 0$, then

$$1\mathbf{u} = 1(u_1, u_2) = (u_1, 0) \neq \mathbf{u}$$

Thus, V is not a vector space with the stated operations.

Our final example will be an unusual vector space that we have included to illustrate how varied vector spaces can be. Since the vectors in this space will be real numbers, it will be important for you to keep track of which operations are intended as vector operations and which ones as ordinary operations on real numbers.

EXAMPLE 8 | An Unusual Vector Space

Let V be the set of positive real numbers, let $\mathbf{u} = u$ and $\mathbf{v} = v$ be any vectors (i.e., positive real numbers) in V, and let k be any scalar. Define the operations on V to be

$$u + v = uv \quad \textbf{[Vector addition is numerical multiplication.]}$$
$$ku = u^k \quad \textbf{[Scalar multiplication is numerical exponentiation.]}$$

Thus, for example, $1 + 1 = 1$ and $(2)(1) = 1^2 = 1$—strange indeed, but nevertheless the set V with these operations satisfies the ten vector space axioms and hence is a vector space. We will confirm Axioms 4, 5, and 7, and leave the others as exercises.

- **Axiom 4**—The zero vector in this space is the number 1 (i.e., $\mathbf{0} = 1$) since

$$u + 1 = u \cdot 1 = u$$

- **Axiom 5**—The negative of a vector u is its reciprocal (i.e., $-u = 1/u$) since

$$u + \frac{1}{u} = u\left(\frac{1}{u}\right) = 1\,(= \mathbf{0})$$

- **Axiom 7**—$k(u + v) = (uv)^k = u^k v^k = (ku) + (kv)$.

Some Properties of Vectors

The following is our first theorem about vector spaces. Although the statements in this theorem closely parallel familiar results in the arithmetic of real numbers, this is no guarantee that they are also true in vector arithmetic, so proof of their validity is required. The proofs are very formal with each step being justified by a vector space axiom or a known property of real numbers. There will not be many rigidly formal proofs of this type in the text, but we have included this one to reinforce the idea that the familiar properties of vectors can all be derived from the vector space axioms.

Theorem 4.1.1

Let V be a vector space, $\mathbf{u}$ a vector in V, and k a scalar; then:

(a) $0\mathbf{u} = \mathbf{0}$

(b) $k\mathbf{0} = \mathbf{0}$

(c) $(-1)\mathbf{u} = -\mathbf{u}$

(d) If $k\mathbf{u} = \mathbf{0}$, then $k = 0$ or $\mathbf{u} = \mathbf{0}$.

We will prove parts (a) and (c) and leave proofs of the remaining parts as exercises.

Proof (a) We can write

$$\begin{aligned} 0\mathbf{u} + 0\mathbf{u} &= (0 + 0)\mathbf{u} && \textbf{[Axiom 8]} \\ &= 0\mathbf{u} && \textbf{[Property of the number 0]} \end{aligned}$$

By Axiom 5 the vector $0\mathbf{u}$ has a negative, $-0\mathbf{u}$. Adding this negative to both sides above yields

$$[0\mathbf{u} + 0\mathbf{u}] + (-0\mathbf{u}) = 0\mathbf{u} + (-0\mathbf{u})$$

or

$$\begin{aligned} 0\mathbf{u} + [0\mathbf{u} + (-0\mathbf{u})] &= 0\mathbf{u} + (-0\mathbf{u}) && \textbf{[Axiom 3]} \\ 0\mathbf{u} + \mathbf{0} &= \mathbf{0} && \textbf{[Axiom 5]} \\ 0\mathbf{u} &= \mathbf{0} && \textbf{[Axiom 4]} \end{aligned}$$

Proof (c) To prove that $(-1)\mathbf{u} = -\mathbf{u}$, we must show that $\mathbf{u} + (-1)\mathbf{u} = \mathbf{0}$. The proof is as follows:

$$\begin{aligned} \mathbf{u} + (-1)\mathbf{u} &= 1\mathbf{u} + (-1)\mathbf{u} && \textbf{[Axiom 10]} \\ &= \big(1 + (-1)\big)\mathbf{u} && \textbf{[Axiom 8]} \\ &= 0\mathbf{u} && \textbf{[Property of numbers]} \\ &= \mathbf{0} && \textbf{[Part (}a\textbf{) of this theorem]} \ \blacksquare \end{aligned}$$

A Closing Observation

This section of the text is important to the overall plan of linear algebra in that it establishes a common thread among such diverse mathematical objects as geometric vectors, vectors in R^n, infinite sequences, matrices, and real-valued functions, to name a few. As a result, whenever we discover a new theorem about general vector spaces, we will at the same time be discovering a theorem about geometric vectors, vectors in R^n, sequences, matrices, real-valued functions, and about any new kinds of vectors that we might discover.

To illustrate this idea, consider what the rather innocent-looking result in part (a) of Theorem 4.1.1 says about the vector space in Example 8. Keeping in mind that the vectors in that space are positive real numbers, that scalar multiplication means numerical exponentiation, and that the zero vector is the number 1, the equation

$$0\mathbf{u} = \mathbf{0}$$

is really a statement of the familiar fact that if u is a positive real number, then

$$u^0 = 1$$

Exercise Set 4.1

1. Let V be the set of all ordered pairs of real numbers, and consider the following addition and scalar multiplication operations on $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$:

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2), \quad k\mathbf{u} = (0, ku_2)$$

- **a.** Compute $\mathbf{u} + \mathbf{v}$ and $k\mathbf{u}$ for $\mathbf{u} = (-1, 2)$, $\mathbf{v} = (3, 4)$, and $k = 3$.
- **b.** In words, explain why V is closed under addition and scalar multiplication.
- **c.** Since addition on V is the standard addition operation on R^2, certain vector space axioms hold for V because they are known to hold for R^2. Which axioms are they?
- **d.** Show that Axioms 7, 8, and 9 hold.
- **e.** Show that Axiom 10 fails and hence that V is not a vector space under the given operations.

2. Let V be the set of all ordered pairs of real numbers, and consider the following addition and scalar multiplication operations on $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$:

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1 + 1, u_2 + v_2 + 1), \quad k\mathbf{u} = (ku_1, ku_2)$$

- **a.** Compute $\mathbf{u} + \mathbf{v}$ and $k\mathbf{u}$ for $\mathbf{u} = (0, 4)$, $\mathbf{v} = (1, -3)$, and $k = 2$.
- **b.** Show that $(0, 0) \neq \mathbf{0}$.
- **c.** Show that $(-1, -1) = \mathbf{0}$.
- **d.** Show that Axiom 5 holds by producing a vector $-\mathbf{u}$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$ for $\mathbf{u} = (u_1, u_2)$.
- **e.** Find two vector space axioms that fail to hold.

In Exercises 3–12, determine whether each set equipped with the given operations is a vector space. For those that are not vector spaces identify the vector space axioms that fail.

3. The set of all real numbers with the standard operations of addition and multiplication.

4. The set of all pairs of real numbers of the form $(x, 0)$ with the standard operations on R^2.

5. The set of all pairs of real numbers of the form (x, y), where $x \geq 0$, with the standard operations on R^2.

6. The set of all n-tuples of real numbers that have the form $(x, x, \ldots, x)$ with the standard operations on R^n.

7. The set of all triples of real numbers with the standard vector addition but with scalar multiplication defined by

$$k(x, y, z) = (k^2x, k^2y, k^2z)$$

8. The set of all 2×2 invertible matrices with the standard matrix addition and scalar multiplication.

9. The set of all 2×2 matrices of the form

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$

with the standard matrix addition and scalar multiplication.

10. The set of all real-valued functions f defined everywhere on the real line and such that $f(1) = 0$ with the operations used in Example 6.

11. The set of all pairs of real numbers of the form $(1, x)$ with the operations

$$(1, y) + (1, y') = (1, y + y') \quad \text{and} \quad k(1, y) = (1, ky)$$

12. The set of polynomials of the form $a_0 + a_1 x$ with the operations

$$(a_0 + a_1 x) + (b_0 + b_1 x) = (a_0 + b_0) + (a_1 + b_1)x$$

and

$$k(a_0 + a_1 x) = (ka_0) + (ka_1)x$$

13. Verify Axioms 3, 7, 8, and 9 for the vector space given in Example 4.

14. Verify Axioms 1, 2, 3, 7, 8, 9, and 10 for the vector space given in Example 6.

15. With the addition and scalar multiplication operations defined in Example 7, show that $V = R^2$ satisfies Axioms 1–9.

16. Verify Axioms 1, 2, 3, 6, 8, 9, and 10 for the vector space given in Example 8.

17. Show that the set of all points in R^2 lying on a line is a vector space with respect to the standard operations of vector addition and scalar multiplication if and only if the line passes through the origin.

18. Show that the set of all points in R^3 lying in a plane is a vector space with respect to the standard operations of vector addition and scalar multiplication if and only if the plane passes through the origin.

In Exercises 19–20, let V be the vector space of positive real numbers with the vector space operations given in Example 8. Let $\mathbf{u} = u$ be any vector in V, and rewrite the vector statement as a statement about real numbers.

19. $-\mathbf{u} = (-1)\mathbf{u}$

20. $k\mathbf{u} = \mathbf{0}$ if and only if $k = 0$ or $\mathbf{u} = \mathbf{0}$.

Working with Proofs

21. The argument that follows proves that if $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ are vectors in a vector space V such that $\mathbf{u} + \mathbf{w} = \mathbf{v} + \mathbf{w}$, then $\mathbf{u} = \mathbf{v}$ (the ***cancellation law*** for vector addition). As illustrated, justify the steps by filling in the blanks.

$\mathbf{u} + \mathbf{w} = \mathbf{v} + \mathbf{w}$	Hypothesis
$(\mathbf{u} + \mathbf{w}) + (-\mathbf{w}) = (\mathbf{v} + \mathbf{w}) + (-\mathbf{w})$	Add $-\mathbf{w}$ to both sides.
$\mathbf{u} + [\mathbf{w} + (-\mathbf{w})] = \mathbf{v} + [\mathbf{w} + (-\mathbf{w})]$	________
$\mathbf{u} + \mathbf{0} = \mathbf{v} + \mathbf{0}$	________
$\mathbf{u} = \mathbf{v}$	________

22. The seven-step proof of part (*b*) of Theorem 4.1.1 follows. Justify each step either by stating that it is true by *hypothesis* or by specifying which of the ten vector space axioms applies.

Hypothesis: Let $\mathbf{u}$ be any vector in a vector space V, let $\mathbf{0}$ be the zero vector in V, and let k be a scalar.

Conclusion: Then $k\mathbf{0} = \mathbf{0}$.

Proof: (1) $k\mathbf{0} + k\mathbf{u} = k(\mathbf{0} + \mathbf{u})$

(2) $= k\mathbf{u}$

(3) Since $k\mathbf{u}$ is in V, $-k\mathbf{u}$ is in V.

(4) Therefore, $(k\mathbf{0} + k\mathbf{u}) + (-k\mathbf{u}) = k\mathbf{u} + (-k\mathbf{u})$.

(5) $k\mathbf{0} + (k\mathbf{u} + (-k\mathbf{u})) = k\mathbf{u} + (-k\mathbf{u})$

(6) $k\mathbf{0} + \mathbf{0} = \mathbf{0}$

(7) $k\mathbf{0} = \mathbf{0}$

In Exercises 23–24, let $\mathbf{u}$ be any vector in a vector space V. Give a step-by-step proof of the stated result using Exercises 21 and 22 as models for your presentation.

23. $0\mathbf{u} = \mathbf{0}$

24. $-\mathbf{u} = (-1)\mathbf{u}$

In Exercises 25–27, prove that the given set with the stated operations is a vector space.

25. The set $V = \{\mathbf{0}\}$ with the operations of addition and scalar multiplication given in Example 1.

26. The set R^∞ of all infinite sequences of real numbers with the operations of addition and scalar multiplication given in Example 3.

27. The set M_{mn} of all $m \times n$ matrices with the usual operations of addition and scalar multiplication.

28. Prove: If $\mathbf{u}$ is a vector in a vector space V and k a scalar such that $k\mathbf{u} = \mathbf{0}$, then either $k = 0$ or $\mathbf{u} = \mathbf{0}$. [*Suggestion:* Show that if $k\mathbf{u} = \mathbf{0}$ and $k \neq 0$, then $\mathbf{u} = \mathbf{0}$. The result then follows as a logical consequence of this.]

True-False Exercises

TF. In parts (a)–(f) determine whether the statement is true or false, and justify your answer.

a. A vector is any element of a vector space.

b. A vector space must contain at least two vectors.

c. If $\mathbf{u}$ is a vector and k is a scalar such that $k\mathbf{u} = \mathbf{0}$, then it must be true that $k = 0$.

d. The set of positive real numbers is a vector space if vector addition and scalar multiplication are the usual operations of addition and multiplication of real numbers.

e. In every vector space the vectors $(-1)\mathbf{u}$ and $-\mathbf{u}$ are the same.

f. In the vector space $F(-\infty, \infty)$ any function whose graph passes through the origin is a zero vector.

4.2 Subspaces

It is often the case that some vector space of interest is contained within a larger vector space whose properties are known. In this section we will show how to recognize when this is the case, we will explain how the properties of the larger vector space can be used to obtain properties of the smaller vector space, and we will give a variety of important examples.

We begin with some terminology.

Definition 1

A subset W of a vector space V is called a ***subspace*** of V if W is itself a vector space under the addition and scalar multiplication defined on V.

In general, to show that a nonempty set W with two operations is a vector space one must verify the ten vector space axioms. However, if W is a subspace of a known vector space V, then certain axioms need not be verified because they are "inherited" from V. For example, it is *not* necessary to verify that $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$ holds in W because it holds for all vectors in V including those in W. On the other hand, it *is* necessary to verify that W is closed under addition and scalar multiplication since it is possible that adding two vectors in W or multiplying a vector in W by a scalar produces a vector in V that is outside of W (**Figure 4.2.1**). Those axioms that are *not* inherited by W are

Axiom 1—Closure of W under addition

Axiom 4—Existence of a zero vector in W

Axiom 5—Existence of a negative in W for every vector in W

Axiom 6—Closure of W under scalar multiplication

so these must be verified to prove that it is a subspace of V. However, the next theorem shows that if Axiom 1 and Axiom 6 hold in W, then Axioms 4 and 5 hold in W as a consequence and hence need not be verified.

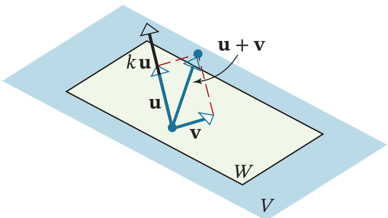


FIGURE 4.2.1 The vectors $\mathbf{u}$ and $\mathbf{v}$ are in W, but the vectors $\mathbf{u}+\mathbf{v}$ and $k\mathbf{u}$ are not.

Theorem 4.2.1

Subspace Test

If W is a nonempty set of vectors in a vector space V, then W is a subspace of V if and only if the following conditions are satisfied.

(*a*) If $\mathbf{u}$ and $\mathbf{v}$ are vectors in W, then $\mathbf{u}+\mathbf{v}$ is in W.

(*b*) If k is a scalar and $\mathbf{u}$ is a vector in W, then $k\mathbf{u}$ is in W.

The Subspace Test states that W is a subspace of V if and only if it is closed under addition and scalar multiplication.

Proof If W is a subspace of V, then all the vector space axioms hold in W, including Axioms 1 and 6, which are precisely conditions (a) and (b).

Conversely, assume that conditions (a) and (b) hold. Since these are Axioms 1 and 6, and since Axioms 2, 3, 7, 8, 9, and 10 are inherited from V, we only need to show that Axioms 4 and 5 hold in W. For this purpose, let $\mathbf{u}$ be any vector in W. It follows from condition (b) that the product $k\mathbf{u}$ is also a vector in W for every scalar k. In particular, $0\mathbf{u} = \mathbf{0}$ and $(-1)\mathbf{u} = -\mathbf{u}$ are in W, which shows that Axioms 4 and 5 hold in W. ■

It is important to note that the first step in applying the Subspace Test to a set W is to confirm that the set is nonempty. This should be clear for all of the examples in this section, so we will omit its explicit verification.

EXAMPLE 1 | The Zero Subspace

If V is any vector space, and if $W = \{\mathbf{0}\}$ is the subset of V that consists of the zero vector only, then W is closed under addition and scalar multiplication since

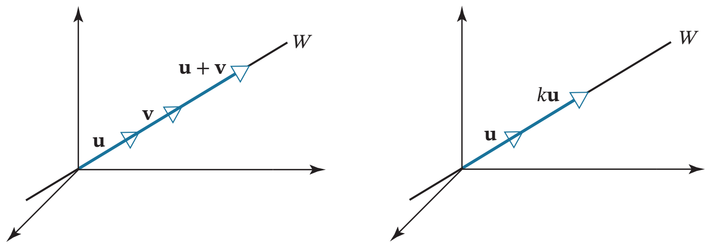
$$\mathbf{0} + \mathbf{0} = \mathbf{0} \quad \text{and} \quad k\mathbf{0} = \mathbf{0}$$

for any scalar k. We call W the ***zero subspace*** of V.

Note that every vector space has at least two subspaces, itself and its zero subspace.

EXAMPLE 2 | Lines Through the Origin Are Subspaces of R^2 and of R^3

If W is a line through the origin of either R^2 or R^3, then adding two vectors on the line or multiplying a vector on the line by a scalar produces another vector on the line, so W is closed under addition and scalar multiplication (see **Figure 4.2.2** for an illustration in R^3).



(a) W is closed under addition.

(b) W is closed under scalar multiplication.

FIGURE 4.2.2

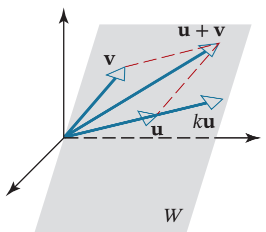


FIGURE 4.2.3 The vectors $\mathbf{u} + \mathbf{v}$ and $k\mathbf{u}$ both lie in the same plane as $\mathbf{u}$ and $\mathbf{v}$.

EXAMPLE 3 | Planes Through the Origin Are Subspaces of R^3

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in a plane W through the origin of R^3, then it is evident geometrically that $\mathbf{u} + \mathbf{v}$ and $k\mathbf{u}$ also lie in the same plane W for any scalar k (**Figure 4.2.3**). Thus W is closed under addition and scalar multiplication.

Table 1 gives a list of subspaces of R^2 and of R^3 that we have encountered thus far. We will see later that these are the *only* subspaces of R^2 and of R^3.

TABLE 1

Subspaces of R^2	Subspaces of R^3
• $\{\mathbf{0}\}$ • Lines through the origin • R^2	• $\{\mathbf{0}\}$ • Lines through the origin • Planes through the origin • R^3

EXAMPLE 4 | A Subset of R^2 That Is Not a Subspace

Let W be the set of all points (x, y) in R^2 for which $x \geq 0$ and $y \geq 0$ (the shaded region in **Figure 4.2.4**). This set is not a subspace of R^2 because it is not closed under scalar multiplication. For example, $\mathbf{v} = (1, 1)$ is a vector in W, but $(-1)\mathbf{v} = (-1, -1)$ is not.

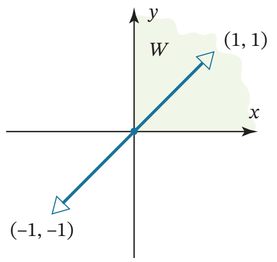


FIGURE 4.2.4 W is not closed under scalar multiplication.

EXAMPLE 5 | Subspaces of M_{nn}

We know from Theorem 1.7.2 that the sum of two symmetric $n \times n$ matrices is symmetric and that a scalar multiple of a symmetric $n \times n$ matrix is symmetric. Thus, the set of symmetric $n \times n$ matrices is closed under addition and scalar multiplication and hence is a subspace of M_{nn}. Similarly, the sets of upper triangular matrices, lower triangular matrices, and diagonal matrices are subspaces of M_{nn}.

EXAMPLE 6 | A Subset of M_{nn} That Is Not a Subspace

The set W of invertible $n \times n$ matrices is not a subspace of M_{nn}, failing on two counts—it is not closed under addition and not closed under scalar multiplication. We will illustrate this with an example in M_{22} that you can readily adapt to M_{nn}. Consider the matrices

$$U = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} -1 & 2 \\ -2 & 5 \end{bmatrix}$$

The matrix $0U$ is the 2×2 zero matrix and hence is not invertible, and the matrix $U + V$ has a column of zeros so it also is not invertible.

EXAMPLE 7 | The Subspace $C(-\infty, \infty)$

CALCULUS REQUIRED

There is a theorem in calculus which states that a sum of continuous functions is continuous and that a constant times a continuous function is continuous. Rephrased in vector language, the set of continuous functions on $(-\infty, \infty)$ is a subspace of $F(-\infty, \infty)$. We will denote this subspace by $C(-\infty, \infty)$.

CALCULUS REQUIRED

EXAMPLE 8 | Functions with Continuous Derivatives

A function with a continuous derivative is said to be *continuously differentiable*. There is a theorem in calculus which states that the sum of two continuously differentiable functions is continuously differentiable and that a constant times a continuously differentiable function is continuously differentiable. Thus, the functions that are continuously differentiable on $(-\infty, \infty)$ form a subspace of $F(-\infty, \infty)$. We will denote this subspace by $C^1(-\infty, \infty)$, where the superscript emphasizes that the *first* derivatives are continuous. To take this a step further, the set of functions with m continuous derivatives on $(-\infty, \infty)$ is a subspace of $F(-\infty, \infty)$ as is the set of functions with derivatives of all orders on $(-\infty, \infty)$. We will denote these subspaces by $C^m(-\infty, \infty)$ and $C^\infty(-\infty, \infty)$, respectively.

EXAMPLE 9 | The Subspace of All Polynomials

Recall that a ***polynomial*** is a function that can be expressed in the form

$$p(x) = a_0 + a_1x + \cdots + a_nx^n \tag{1}$$

where $a_0, a_1, \ldots, a_n$ are constants. It is evident that the sum of two polynomials is a polynomial and that a constant times a polynomial is a polynomial. Thus, the set W of all polynomials is closed under addition and scalar multiplication and hence is a subspace of $F(-\infty, \infty)$. We will denote this space by P_∞.

EXAMPLE 10 | The Subspace of Polynomials of Degree $\leq n$

Recall that the ***degree*** of a polynomial is the highest power of the variable that occurs with a nonzero coefficient. Thus, for example, if $a_n \neq 0$ in Formula (1), then that polynomial has degree n. It is *not* true that the set W of polynomials with positive degree n is a subspace of $F(-\infty, \infty)$ because that set is not closed under addition. For example, the polynomials

$$1 + 2x + 3x^2 \quad \text{and} \quad 5 + 7x - 3x^2$$

both have degree 2, but their sum has degree 1. What *is* true, however, is that for each nonnegative integer n the polynomials of degree n *or less* form a subspace of $F(-\infty, \infty)$. We will denote this space by P_n.

In this text we regard all constants to be polynomials of degree zero. Be aware, however, that some authors do not assign a degree to the constant 0.

The Hierarchy of Function Spaces

It is proved in calculus that polynomials are continuous functions and have continuous derivatives of all orders on $(-\infty, \infty)$. Thus, it follows that P_∞ is not only a subspace of $F(-\infty, \infty)$, as previously observed, but is also a subspace of $C^\infty(-\infty, \infty)$. We leave it for you to convince yourself that the vector spaces discussed in Examples 7 to 10 are "nested" one inside the other as illustrated in **Figure 4.2.5**.

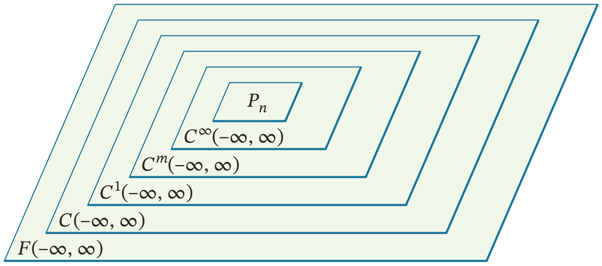


FIGURE 4.2.5

Remark In our previous examples we considered functions that were defined at all points of the interval $(-\infty, \infty)$. Sometimes we will want to consider functions that are only defined on some subinterval of $(-\infty, \infty)$, say the closed interval $[a, b]$ or the open interval (a, b). In such cases we will make an appropriate notation change. For example, $C[a, b]$ is the space of continuous functions on $[a, b]$ and $C(a, b)$ is the space of continuous functions on (a, b).

In the following examples we will illustrate how the Subspace Test can be applied to various nonempty subsets of R^n, M_{mn}, P_n, and $F(-\infty, \infty)$.

EXAMPLE 11 | Applying the Subspace Test in M_{22}

Determine whether the indicated set of matrices is a subspace of M_{22}.

(*a*) The set U consisting of all matrices of the form

$$\begin{bmatrix} x & 0 \\ 2x & y \end{bmatrix} \tag{2}$$

(*b*) The set W consisting of all 2×2 matrices A such that

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

Solution (*a*) If A and B are matrices in U, then they can be expressed in the form

$$A = \begin{bmatrix} a & 0 \\ 2a & b \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} c & 0 \\ 2c & d \end{bmatrix}$$

for some real numbers a, b, c, and d. But

$$A + B = \begin{bmatrix} a + c & 0 \\ 2(a + c) & b + d \end{bmatrix}$$

is also a matrix in U since it is of form (2) with $x = a + c$ and $y = b + d$. Thus, U is closed under addition. Similarly, U is closed under scalar multiplication since

$$kA = \begin{bmatrix} ka & 0 \\ 2ka & kb \end{bmatrix}$$

is of form (2) with $x = ka$ and $y = kb$. These two results establish that U is a subspace of M_{22}.

Solution (*b*) The set W is not a subspace of M_{22}. To see that this is so, it suffices to show that W is *either* not closed under addition or not closed under scalar multiplication. To see that it is not closed under scalar multiplication, let

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

This is a vector in W since

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

so A satisfies Equation (3). However, $2A$ does not satisfy Equation (3) since

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

and hence is not a vector in W. This alone establishes that W is not a subspace of M_{22}. However, it is also true that W is not closed under addition. We leave the proof for the reader.

EXAMPLE 12 | Applying the Subspace Test in P_2

Determine whether the indicated set of polynomials is a subspace of P_2.

(*a*) The set U consisting of all polynomials of the form $\mathbf{p} = 1 + ax - ax^2$, where a is a real number.

(*b*) The set W consisting of all polynomials $\mathbf{p}$ in P_2 such that $\mathbf{p}(2) = 0$.

Solution (*a*) The set U is not a subspace of P_2 because it is not closed under addition. For example, the polynomials $\mathbf{p} = 1 + x - x^2$ and $\mathbf{q} = 1 + 2x - 2x^2$ are in U, but

$$\mathbf{p} + \mathbf{q} = 2 + 3x - 3x^2$$

is not. We leave it for you to verify that U is also not closed under scalar multiplication.

Solution (*b*) If $\mathbf{p}$ and $\mathbf{q}$ are polynomials in W, and k is any real number, then

$$(\mathbf{p} + \mathbf{q})(2) = \mathbf{p}(2) + \mathbf{q}(2) = 0 + 0 = 0$$

and

$$(k\mathbf{p})(2) = k \cdot \mathbf{p}(2) = k \cdot 0 = 0.$$

Since $\mathbf{p} + \mathbf{q}$ and $k\mathbf{p}$ are in W, it follows that W is a subspace of P_2.

Building Subspaces

The following theorem provides a useful way of creating a new subspace from known subspaces.

Theorem 4.2.2

If $W_1, W_2, \ldots, W_r$ are subspaces of a vector space V, then the intersection of these subspaces is also a subspace of V.

Note that the first step in proving Theorem 4.2.2 was to establish that W contained at least one vector. This is important, for otherwise the subsequent argument might be logically correct but meaningless.

Proof Let W be the intersection of the subspaces $W_1, W_2, \ldots, W_r$. This set is not empty because each of these subspaces contains the zero vector of V, and hence so does their intersection. Thus, it remains to show that W is closed under addition and scalar multiplication.

To prove closure under addition, let $\mathbf{u}$ and $\mathbf{v}$ be vectors in W. Since W is the intersection of $W_1, W_2, \ldots, W_r$, it follows that $\mathbf{u}$ and $\mathbf{v}$ also lie in each of these subspaces. Moreover, since these subspaces are closed under addition and scalar multiplication, they also all contain the vectors $\mathbf{u} + \mathbf{v}$ and $k\mathbf{u}$ for every scalar k, and hence so does their intersection W. This proves that W is closed under addition and scalar multiplication. ■

Solution Spaces of Homogeneous Systems

The solutions of a homogeneous linear system $A\mathbf{x} = \mathbf{0}$ of m equations in n unknowns can be viewed as vectors in R^n. The following theorem provides an important insight into the geometric structure of the solution set.

Theorem 4.2.3

The solution set of a homogeneous system $A\mathbf{x} = \mathbf{0}$ of m equations in n unknowns is a subspace of R^n.

Proof Let W be the solution set of the system. The set W is not empty because it contains at least the trivial solution $\mathbf{x} = \mathbf{0}$.

To show that W is a subspace of R^n, we must show that it is closed under addition and scalar multiplication. To do this, let $\mathbf{x}_1$ and $\mathbf{x}_2$ be vectors in W. Since these vectors are solutions of $A\mathbf{x} = \mathbf{0}$, we have

$$A\mathbf{x}_1 = \mathbf{0} \quad \text{and} \quad A\mathbf{x}_2 = \mathbf{0}$$

It follows from these equations and the distributive property of matrix multiplication that

$$A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

so W is closed under addition. Similarly, if k is any scalar then

$$A(k\mathbf{x}_1) = kA\mathbf{x}_1 = k\mathbf{0} = \mathbf{0}$$

so W is also closed under scalar multiplication. ■

Because the solution set of a homogeneous system in n unknowns is actually a subspace of R^n, we will generally refer to it as the ***solution space*** of the system.

EXAMPLE 13 | Solution Spaces of Homogeneous Systems

In each part the solution of the linear system is provided. Give a geometric description of the solution set.

$$(a)\ \begin{bmatrix} 1 & -2 & 3 \\ 2 & -4 & 6 \\ 3 & -6 & 9 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad (b)\ \begin{bmatrix} 1 & -2 & 3 \\ -3 & 7 & -8 \\ -2 & 4 & -6 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$(c)\ \begin{bmatrix} 1 & -2 & 3 \\ -3 & 7 & -8 \\ 4 & 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad (d)\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Solution (a) The solutions are

$$x = 2s - 3t, \quad y = s, \quad z = t$$

from which it follows that

$$x = 2y - 3z \quad \text{or} \quad x - 2y + 3z = 0$$

This is the equation of a plane through the origin that has $\mathbf{n} = (1, -2, 3)$ as a normal.

Solution (b) The solutions are

$$x = -5t, \quad y = -t, \quad z = t$$

which are parametric equations for the line through the origin that is parallel to the vector $\mathbf{v} = (-5, -1, 1)$.

Solution (c) The only solution is $x = 0$, $y = 0$, $z = 0$, so the solution space consists of the single point $\{\mathbf{0}\}$.

Solution (d) This linear system is satisfied by all real values of x, y, and z, so the solution space is all of R^3.

Remark Whereas the solution set of every *homogeneous* system of m equations in n unknowns is a subspace of R^n, it is *never* true that the solution set of a *nonhomogeneous* system of m equations in n unknowns is a subspace of R^n. There are two possible scenarios: first, the system may not have any solutions at all, and second, if there are solutions, then the solution set will not be closed under either addition or scalar multiplication (Exercise 22).

The Linear Transformation Viewpoint

Theorem 4.2.3 can be viewed as a statement about matrix transformations by letting $T_A : R^n \to R^m$ be multiplication by the coefficient matrix A. From this point of view the solution space of $A\mathbf{x} = \mathbf{0}$ is the set of vectors in R^n that T_A maps into the zero vector in R^m. This set is sometimes called the ***kernel*** of the transformation, so with this terminology Theorem 4.2.3 can be rephrased as follows.

Theorem 4.2.4

If A is an $m \times n$ matrix, then the kernel of the matrix transformation $T_A : R^n \to R^m$ is a subspace of R^n.

Exercise Set 4.2

In Exercises 1–2, use the Subspace Test to determine which of the sets are subspaces of R^3.

1. **a.** All vectors of the form $(a, 0, 0)$.

   b. All vectors of the form $(a, 1, 1)$.

   c. All vectors of the form (a, b, c), where $b = a + c$.

2. **a.** All vectors of the form (a, b, c), where $b = a + c + 1$.

   b. All vectors of the form $(a, b, 0)$.

   c. All vectors of the form (a, b, c) for which $a + b = 7$.

In Exercises 3–4, use the Subspace Test to determine which of the sets are subspaces of M_{nn}.

3. **a.** The set of all diagonal $n \times n$ matrices.

   b. The set of all $n \times n$ matrices A such that $\det(A) = 0$.

   c. The set of all $n \times n$ matrices A such that $\operatorname{tr}(A) = 0$.

   d. The set of all symmetric $n \times n$ matrices.

4. **a.** The set of all $n \times n$ matrices A such that $A^T = -A$.

   b. The set of all $n \times n$ matrices A for which $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

   c. The set of all $n \times n$ matrices A such that $AB = BA$ for some fixed $n \times n$ matrix B.

   d. The set of all invertible $n \times n$ matrices.

In Exercises 5–6, use the Subspace Test to determine which of the sets are subspaces of P_3.

5. **a.** All polynomials $a_0 + a_1x + a_2x^2 + a_3x^3$ for which $a_0 = 0$.

   b. All polynomials $a_0 + a_1x + a_2x^2 + a_3x^3$ for which $a_0 + a_1 + a_2 + a_3 = 0$.

6. **a.** All polynomials of the form $a_0 + a_1x + a_2x^2 + a_3x^3$ in which a_0, a_1, a_2, and a_3 are rational numbers.

   b. All polynomials of the form $a_0 + a_1x$, where a_0 and a_1 are real numbers.

In Exercises 7–8, use the Subspace Test to determine which of the sets are subspaces of $F(-\infty, \infty)$.

7. **a.** All functions f in $F(-\infty, \infty)$ for which $f(0) = 0$.

b. All functions f in $F(-\infty, \infty)$ for which $f(0) = 1$.

8. **a.** All functions f in $F(-\infty, \infty)$ for which $f(-x) = f(x)$.

b. All polynomials of degree 2.

In Exercises 9–10, use the Subspace Test to determine which of the sets are subspaces of R^∞.

9. **a.** All sequences $\mathbf{v}$ in R^∞ of the form $\mathbf{v} = (v, 0, v, 0, v, 0, \ldots)$.

b. All sequences $\mathbf{v}$ in R^∞ of the form $\mathbf{v} = (v, 1, v, 1, v, 1, \ldots)$.

10. **a.** All sequences $\mathbf{v}$ in R^∞ of the form

$$\mathbf{v} = (v, 2v, 4v, 8v, 16v, \ldots)$$

b. All sequences in R^∞ whose components are 0 from some point on.

In Exercises 11–12, use the Subspace Test to determine which of the sets are subspaces of M_{22}.

11. **a.** All matrices of the form $\begin{bmatrix} a & 0 \\ b & 0 \end{bmatrix}$.

b. All matrices of the form $\begin{bmatrix} a & 1 \\ b & 1 \end{bmatrix}$.

c. All 2×2 matrices A such that

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

12. **a.** All 2×2 matrices A such that

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

b. All 2×2 matrices A such that

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

c. All 2×2 matrices A for which $\det(A) = 0$.

In Exercises 13–14, use the Subspace Test to determine which of the sets are subspaces of R^4.

13. **a.** All vectors of the form (a, a^2, a^3, a^4).

b. All vectors of the form $(a, 0, b, 0)$.

14. **a.** All vectors $\mathbf{x}$ in R^4 such that $A\mathbf{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, where

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

b. All vectors $\mathbf{x}$ in R^4 such that $A\mathbf{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, where A is as in part (a).

In Exercises 15–16, use the Subspace Test to determine which of the sets are subspaces of P_∞.

15. **a.** All polynomials of degree less than or equal to 6.

b. All polynomials of degree equal to 6.

c. All polynomials of degree greater than or equal to 6.

16. **a.** All polynomials with even coefficients.

b. All polynomials whose coefficients sum to 0.

c. All polynomials of even degree.

17. **(*Calculus Required*)** Which of the following are subspaces of R^∞?

a. All sequences of the form $\mathbf{v} = (v_1, v_2, \ldots, v_n, \ldots)$ such that $\lim_{n\to\infty} v_n = 0$.

b. All convergent sequences (that is, all sequences of the form $\mathbf{v} = (v_1, v_2, \ldots, v_n, \ldots)$ such that $\lim_{n\to\infty} v_n$ exists).

c. All sequences of the form $\mathbf{v} = (v_1, v_2, \ldots, v_n, \ldots)$ such that $\sum_{n=1}^{\infty} v_n = 0$.

d. All sequences of the form $\mathbf{v} = (v_1, v_2, \ldots, v_n, \ldots)$ such that $\sum_{n=1}^{\infty} v_n$ converges.

18. A line L through the origin in R^3 can be represented by parametric equations of the form $x = at$, $y = bt$, and $z = ct$. Use these equations to show that L is a subspace of R^3 by showing that if $\mathbf{v}_1 = (x_1, y_1, z_1)$ and $\mathbf{v}_2 = (x_2, y_2, z_2)$ are points on L and k is any real number, then $k\mathbf{v}_1$ and $\mathbf{v}_1 + \mathbf{v}_2$ are also points on L.

19. Determine whether the solution space of the system $A\mathbf{x} = \mathbf{0}$ is a line through the origin, a plane through the origin, or the origin only. If it is a plane, find an equation for it. If it is a line, find parametric equations for it.

a. $A = \begin{bmatrix} -1 & 1 & 1 \\ 3 & -1 & 0 \\ 2 & -4 & -5 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}$

c. $A = \begin{bmatrix} 1 & -3 & 1 \\ 2 & -6 & 2 \\ 3 & -9 & 3 \end{bmatrix}$ **d.** $A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & -1 & 4 \\ 3 & 1 & 11 \end{bmatrix}$

20. **(*Calculus required*)** Show that the following sets of functions are subspaces of $F(-\infty, \infty)$.

a. All continuous functions on $(-\infty, \infty)$.

b. All differentiable functions on $(-\infty, \infty)$.

c. All differentiable functions on $(-\infty, \infty)$ that satisfy $\mathbf{f}' + 2\mathbf{f} = \mathbf{0}$.

21. **(*Calculus required*)** Show that the set of continuous functions $\mathbf{f} = f(x)$ on $[a, b]$ such that

$$\int_a^b f(x)\,dx = 0$$

is a subspace of $C[a, b]$.

22. Show that the solution vectors of a consistent nonhomogeneous system of m linear equations in n unknowns do not form a subspace of R^n.

23. If T_A is multiplication by a matrix A with three columns, then the kernel of T_A is one of four possible geometric objects. What are they? Explain how you reached your conclusion.

24. Consider the following subsets of P_3: V consists of all polynomials $a_0 + a_1x + a_2x^2 + a_3x^3$ such that $a_0 + a_3 = 0$ and W consists of all polynomials $\mathbf{p}$ such that $\mathbf{p}(1) = 0$.

a. Use the Subspace Test to show that V and W are subspaces of P_3.

b. Show that the set of all polynomials

$$\mathbf{p} = a_0 + a_1x + a_2x^2 + a_3x^3$$

such that $a_0 + a_3 = 0$ *and* $\mathbf{p}(1) = 0$ is a subspace of P_3 *without* using the Subspace Test.

25. The accompanying figure shows a mass-spring system in which a block of mass m is set into vibratory motion by pulling the block beyond its natural position at $x = 0$ and releasing it at time $t = 0$. If friction and air resistance are ignored, then the x-coordinate $x(t)$ of the block at time t is given by a function of the form

$$x(t) = c_1 \cos \omega t + c_2 \sin \omega t$$

where ω is a fixed constant that depends on the mass of the block and the stiffness of the spring and c_1 and c_2 are arbitrary. Show that this set of functions forms a subspace of $C^\infty(-\infty, \infty)$.

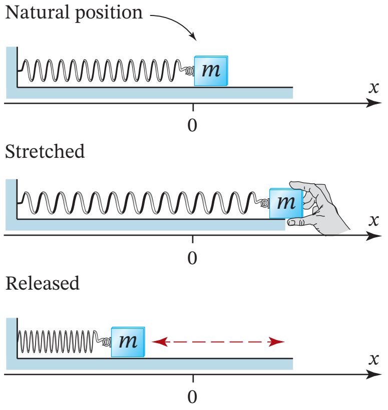


FIGURE Ex-25

26. Show that Theorem 4.2.2 would be false if the word "intersection" was replaced with "union" by giving an example of a vector space V and subspaces U and W such that the union of U with W is not a subspace of V.

Working with Proofs

27. A function $\mathbf{f} = f(x)$ in $F(-\infty, \infty)$ is **even** if $f(-a) = f(a)$ for all real numbers a. Prove that the set of even functions is a subspace of $F(-\infty, \infty)$.

28. A function $\mathbf{f} = f(x)$ in $F(-\infty, \infty)$ is **odd** if $f(-a) = -f(a)$ for all real numbers a. Prove that the set of odd functions is a subspace of $F(-\infty, \infty)$.

29. If U and W are subspaces of a vector space V, then the **sum** of U and W is the set $U + W$ consisting of all vectors of the form $\mathbf{u} + \mathbf{w}$, where $\mathbf{u}$ is a vector in U and $\mathbf{w}$ is a vector in W. Prove that $U + W$ is a subspace of V.

True-False Exercises

TF. In parts **(a)–(h)** determine whether the statement is true or false, and justify your answer.

a. Every subspace of a vector space is itself a vector space.

b. Every vector space is a subspace of itself.

c. Every subset of a vector space V that contains the zero vector in V is a subspace of V.

d. The kernel of a matrix transformation $T_A : R^n \to R^m$ is a subspace of R^m.

e. The solution set of a consistent linear system $A\mathbf{x} = \mathbf{b}$ of m equations in n unknowns is a subspace of R^n.

f. The intersection of any two subspaces of a vector space V is a subspace of V.

g. The union of any two subspaces of a vector space V is a subspace of V.

h. The set of upper triangular $n \times n$ matrices is a subspace of the vector space of all $n \times n$ matrices.

Working with Technology

T1. Determine whether the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are in the kernel of T_A, where

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \\ 21 & 22 & 23 & 24 & 25 \end{bmatrix}$$

and $\mathbf{u} = (1, -2, 1, 0, 0)$, $\mathbf{v} = (5, 0, 1, -2, 1)$, $\mathbf{w} = (3, -4, 0, 0, 1)$

4.3 Spanning Sets

It is often the case that all of the vectors in a vector space V can be expressed in terms of some small subset S of vectors in V. The vectors in S can be viewed as the building blocks for constructing all of the vectors in V. This is important because it makes it possible to deduce properties of an entire vector space V by focusing attention on the small set of vectors in S.

The following definition, which generalizes Definition 4 of Section 3.1, is fundamental to the study of vector spaces.

Definition 1

If $\mathbf{w}$ is a vector in a vector space V, then $\mathbf{w}$ is said to be a ***linear combination*** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ in V if $\mathbf{w}$ can be expressed in the form

$$\mathbf{w} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_r\mathbf{v}_r \tag{1}$$

where $k_1, k_2, \ldots, k_r$ are scalars. These scalars are called the ***coefficients*** of the linear combination.

If $r = 1$, then Equation (1) has the form $\mathbf{w} = k_1\mathbf{v}_1$, in which case the linear combination is just a scalar multiple of $\mathbf{v}_1$.

Theorem 4.3.1

If $S = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\}$ is a nonempty set of vectors in a vector space V, then:

(a) The set W of all possible linear combinations of the vectors in S is a subspace of V.

(b) The set W in part (a) is the "smallest" subspace of V that contains all of the vectors in S in the sense that any other subspace that contains those vectors contains W.

Proof (a) Let W be the set of all possible linear combinations of the vectors in S. We must show that W is closed under addition and scalar multiplication. To prove closure under addition, let

$$\mathbf{u} = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_r\mathbf{w}_r \quad \text{and} \quad \mathbf{v} = k_1\mathbf{w}_1 + k_2\mathbf{w}_2 + \cdots + k_r\mathbf{w}_r$$

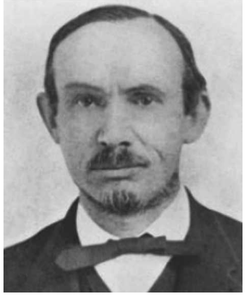
be two vectors in W. It follows that their sum can be written as

$$\mathbf{u} + \mathbf{v} = (c_1 + k_1)\mathbf{w}_1 + (c_2 + k_2)\mathbf{w}_2 + \cdots + (c_r + k_r)\mathbf{w}_r$$

which is a linear combination of the vectors in S. Thus, W is closed under addition. We leave it for you to prove that W is also closed under scalar multiplication and hence is a subspace of V.

Proof (b) Let W' be any subspace of V that contains all of the vectors in S. Since W' is closed under addition and scalar multiplication, it contains all linear combinations of the vectors in S and hence contains W. ■

Historical Note



George William Hill (1838–1914)

The term *linear combination* is due to the American mathematician G. W. Hill, who introduced it in a research paper on planetary motion published in 1900. Hill was a "loner" who preferred to work out of his home in West Nyack, New York, rather than in academia, though he did try lecturing at Columbia University for a few years. Interestingly, he apparently returned the teaching salary, indicating that he did not need the money and did not want to be bothered looking after it. Although technically a mathematician, Hill had little interest in modern developments of mathematics and worked almost entirely on the theory of planetary orbits.

[Image: Courtesy of the American Mathematical Society (www.ams.org)]

In the case where S is the empty set $\emptyset$, it will be convenient to agree that $\text{span}(\emptyset) = \{\mathbf{0}\}$.

The subspace W in Theorem 4.3.1 is called the subspace of V ***spanned*** by S. The vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ in S are said to ***span*** W, and we write

$$W = \text{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\} \quad \text{or} \quad W = \text{span}(S)$$

EXAMPLE 1 | The Standard Unit Vectors Span R^n

Recall that the standard unit vectors in R^n are

$$\mathbf{e}_1 = (1, 0, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, 0, \ldots, 0), \ldots, \quad \mathbf{e}_n = (0, 0, 0, \ldots, 1)$$

These vectors span R^n since every vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ in R^n can be expressed as

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n$$

which is a linear combination of $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. Thus, for example, the vectors

$$\mathbf{i} = (1, 0, 0), \quad \mathbf{j} = (0, 1, 0), \quad \mathbf{k} = (0, 0, 1)$$

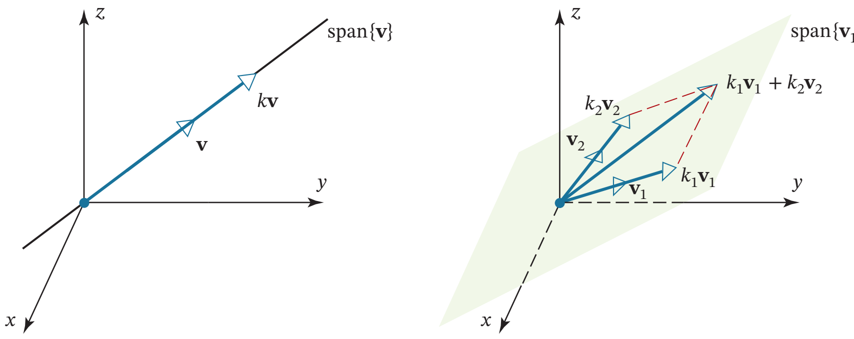
span R^3 since every vector $\mathbf{v} = (a, b, c)$ in this space can be expressed as

$$\mathbf{v} = (a, b, c) = a(1, 0, 0) + b(0, 1, 0) + c(0, 0, 1) = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$$

EXAMPLE 2 | A Geometric View of Spanning in R^2 and R^3

(*a*) If $\mathbf{v}$ is a nonzero vector in R^2 or R^3 that has its initial point at the origin, then $\text{span}\{\mathbf{v}\}$, which is the set of all scalar multiples of $\mathbf{v}$, is the line through the origin determined by $\mathbf{v}$. You should be able to visualize this from **Figure 4.3.1*a*** by observing that the tip of the vector $k\mathbf{v}$ can be made to fall at any point on the line by choosing the value of k to lengthen, shorten, or reverse the direction of $\mathbf{v}$ appropriately.

(*b*) If $\mathbf{v}_1$ and $\mathbf{v}_2$ are nonzero vectors in R^3 that have their initial points at the origin, then $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$, which consists of all linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$, is the plane through the origin determined by these two vectors. You should be able to visualize this from **Figure 4.3.1*b*** by observing that the tip of the vector $k_1\mathbf{v}_1 + k_2\mathbf{v}_2$ can be made to fall at any point in the plane by adjusting the scalars k_1 and k_2 to lengthen, shorten, or reverse the directions of the vectors $k_1\mathbf{v}_1$ and $k_2\mathbf{v}_2$ appropriately.



(***a***) $\text{span}\{\mathbf{v}\}$ is the line through the origin determined by $\mathbf{v}$

(***b***) $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ is the plane through the origin determined by $\mathbf{v}_1$ and $\mathbf{v}_2$

FIGURE 4.3.1

EXAMPLE 3 | A Spanning Set for P_n

The polynomials $1, x, x^2, \ldots, x^n$ span the vector space P_n defined in Example 10 since each polynomial $\mathbf{p}$ in P_n can be written as

$$\mathbf{p} = a_0 + a_1x + \cdots + a_nx^n$$

which is a linear combination of $1, x, x^2, \ldots, x^n$. We can denote this by writing

$$P_n = \operatorname{span}\{1, x, x^2, \ldots, x^n\}$$

The next two examples are concerned with two important types of problems:

- Given a nonempty set S of vectors in R^n and a vector $\mathbf{v}$ in R^n, determine whether $\mathbf{v}$ is a linear combination of the vectors in S.
- Given a nonempty set S of vectors in R^n, determine whether the vectors span R^n.

EXAMPLE 4 | Linear Combinations

Consider the vectors $\mathbf{u} = (1, 2, -1)$ and $\mathbf{v} = (6, 4, 2)$ in R^3. Show that $\mathbf{w} = (9, 2, 7)$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$ and that $\mathbf{w}' = (4, -1, 8)$ is *not* a linear combination of $\mathbf{u}$ and $\mathbf{v}$.

Solution In order for $\mathbf{w}$ to be a linear combination of $\mathbf{u}$ and $\mathbf{v}$, there must be scalars k_1 and k_2 such that $\mathbf{w} = k_1\mathbf{u} + k_2\mathbf{v}$; that is,

$$(9, 2, 7) = k_1(1, 2, -1) + k_2(6, 4, 2) = (k_1 + 6k_2, 2k_1 + 4k_2, -k_1 + 2k_2)$$

Equating corresponding components gives

$$\begin{aligned} k_1 + 6k_2 &= 9 \\ 2k_1 + 4k_2 &= 2 \\ -k_1 + 2k_2 &= 7 \end{aligned}$$

Solving this system using Gaussian elimination yields $k_1 = -3$, $k_2 = 2$, so

$$\mathbf{w} = -3\mathbf{u} + 2\mathbf{v}$$

Similarly, for $\mathbf{w}'$ to be a linear combination of $\mathbf{u}$ and $\mathbf{v}$, there must be scalars k_1 and k_2 such that $\mathbf{w}' = k_1\mathbf{u} + k_2\mathbf{v}$; that is,

$$(4, -1, 8) = k_1(1, 2, -1) + k_2(6, 4, 2) = (k_1 + 6k_2, 2k_1 + 4k_2, -k_1 + 2k_2)$$

Equating corresponding components gives

$$\begin{aligned} k_1 + 6k_2 &= 4 \\ 2k_1 + 4k_2 &= -1 \\ -k_1 + 2k_2 &= 8 \end{aligned}$$

This system of equations is inconsistent (verify), so no such scalars k_1 and k_2 exist. Consequently, $\mathbf{w}'$ is not a linear combination of $\mathbf{u}$ and $\mathbf{v}$.

EXAMPLE 5 | Testing for Spanning

Determine whether the vectors $\mathbf{v}_1 = (1, 1, 2)$, $\mathbf{v}_2 = (1, 0, 1)$, and $\mathbf{v}_3 = (2, 1, 3)$ span the vector space R^3.

Solution We must determine whether an arbitrary vector $\mathbf{b} = (b_1, b_2, b_3)$ in R^3 can be expressed as a linear combination

$$\mathbf{b} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + k_3\mathbf{v}_3$$

of the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. Expressing this equation in terms of components gives

$$(b_1, b_2, b_3) = k_1(1, 1, 2) + k_2(1, 0, 1) + k_3(2, 1, 3)$$

or

$$(b_1, b_2, b_3) = (k_1 + k_2 + 2k_3, k_1 + k_3, 2k_1 + k_2 + 3k_3)$$

or

$$\begin{aligned} k_1 + k_2 + 2k_3 &= b_1 \\ k_1 \qquad + \ k_3 &= b_2 \\ 2k_1 + k_2 + 3k_3 &= b_3 \end{aligned}$$

Thus, our problem reduces to ascertaining whether this system is consistent for all values of b_1, b_2, and b_3. One way of doing this is to use parts (e) and (g) of Theorem 2.3.8, which state that the system is consistent if and only if its coefficient matrix

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

has a nonzero determinant. But this is *not* the case here since $\det(A) = 0$ (verify), so $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ do not span R^3.

In Examples 4 and 5 the question of whether a given set of vectors spans R^3 was answered by determining whether a corresponding linear system was consistent or inconsistent. This suggests a more general procedure for deciding whether a nonempty set of vectors in a vector space V spans V. The procedure we give will be applicable in a wide variety of vector spaces, though later we will encounter vector spaces in which the procedure does not apply and other methods are required.

A Procedure for Identifying Spanning Sets

Step 1. Let $S = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\}$ be a given set of vectors in V, and let $\mathbf{x}$ be an arbitrary vector in V.

Step 2. Set up the augmented matrix for the linear system that results by equating corresponding components on the two sides of the vector equation

$$k_1\mathbf{w}_1 + k_2\mathbf{w}_2 + \cdots + k_r\mathbf{w}_r = \mathbf{x} \tag{2}$$

Step 3. Use the techniques developed in Chapters 1 and 2 to investigate the consistency or inconsistency of that system. If it is consistent for *all* choices of $\mathbf{x}$, the vectors in S span V, and if it is inconsistent for *some* vector $\mathbf{x}$, they do not.

The next two examples illustrate this procedure.

EXAMPLE 6 | Testing for Spanning in P_2

Determine whether the set S spans P_2.

(a) $S = \{1 + x + x^2, -1 - x, 2 + 2x + x^2\}$

(b) $S = \{x + x^2, x - x^2, 1 + x, 1 - x\}$

Solution (a) An arbitrary vector in P_2 is of the form $\mathbf{p} = a + bx + cx^2$, and so (2) becomes

$$k_1(1 + x + x^2) + k_2(-1 - x) + k_3(2 + 2x + x^2) = a + bx + cx^2$$

which we can rewrite as

$$(k_1 - k_2 + 2k_3) + (k_1 - k_2 + 2k_3)x + (k_1 + k_3)x^2 = a + bx + cx^2$$

Equating corresponding coefficients yields a linear system whose augmented matrix is

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

and whose coefficient matrix is

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

Because this matrix is square we can apply Theorem 2.3.8. Since the matrix A has two identical rows it follows that $\det(A) = 0$, so parts (e) and (g) of that theorem imply that the system is inconsistent for *some* choice of a, b, and c; and this tells us that S does *not* span P_2.

Solution (b) Using the same procedure as in part (a), the augmented matrix corresponding to (2) is

$$\begin{bmatrix} 0 & 0 & 1 & -1 & a \\ 1 & 1 & 1 & -1 & b \\ 1 & -1 & 0 & 0 & c \end{bmatrix} \tag{3}$$

Whereas Theorem 2.3.8 was applicable in part (a), it is not applicable here because the coefficient matrix is not square. However, reducing (3) to reduced row echelon form yields (verify)

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \frac{-a+b+c}{2} \\ 0 & 1 & 0 & 0 & \frac{-a+b-c}{2} \\ 0 & 0 & 1 & -1 & a \end{bmatrix}$$

so (3) is consistent for every choice a, b, and c. Thus, the vectors in S span P_2, which we can express by writing $\text{span}(S) = P_2$.

EXAMPLE 7 | Testing for Spanning in M_{22}

In each part, determine whether the set S spans M_{22}.

(a) $S = \left\{ \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right\}$

(b) $S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix} \right\}$

Solution (a) An arbitrary vector in M_{22} is of the form $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, so Equation (2) becomes

$$k_1\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} + k_2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + k_3\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} + k_4\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

which we can rewrite as

$$\begin{bmatrix} k_1 + k_2 + k_3 + k_4 & 2k_1 + 2k_3 + k_4 \\ k_3 + k_4 & k_1 + k_2 + k_4 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Equating corresponding entries produces a linear system whose augmented matrix is

$$\begin{bmatrix} 1 & 1 & 1 & 1 & a \\ 2 & 0 & 2 & 1 & b \\ 0 & 0 & 1 & 1 & c \\ 1 & 1 & 0 & 1 & d \end{bmatrix} \text{ and whose coefficient matrix is } A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 0 & 2 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

As in part (a) of Example 6, the coefficient matrix is square, so we can apply parts (e) and (g) of Theorem 2.3.8. We leave it for you to verify that $\det(A) = -2 \neq 0$, so the system is consistent for *every* choice of a, b, c, and d, which implies that $\text{span}(S) = M_{22}$.

Solution (*b*) Using the same procedure as in part (*a*), the augmented matrix for the linear system corresponding to Equation (2) is

$$\begin{bmatrix} 1 & -1 & 0 & 0 & a \\ 0 & 0 & 0 & 1 & b \\ 0 & 1 & 1 & -1 & c \\ 0 & 0 & 0 & 1 & d \end{bmatrix} \text{ and the coefficient matrix is } A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ which}$$

is square, so once again we can apply parts (*e*) and (*g*) of Theorem 2.3.8. Since the second and fourth rows of this matrix are identical, it follows that $\det(A) = 0$. Thus, the system is inconsistent for *some* choice of a, b, c, and d, which implies that S does not span M_{22}.

A Concluding Observation

It is important to recognize that spanning sets are not unique. For example, any nonzero vector on the line in Figure 4.3.1*a* will span that line, and any two noncollinear vectors in the plane in Figure 4.3.1*b* will span that plane. The following theorem, whose proof is left as an exercise, states conditions under which two sets of vectors will span the same space.

Theorem 4.3.2

If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ and $S' = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ are nonempty sets of vectors in a vector space V, then

$$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} = \operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$$

if and only if each vector in S is a linear combination of those in S', and each vector in S' is a linear combination of those in S.

Exercise Set 4.3

1. Which of the following are linear combinations of $\mathbf{u} = (0, -2, 2)$ and $\mathbf{v} = (1, 3, -1)$?

   a. $(2, 2, 2)$ **b.** $(0, 4, 5)$ **c.** $(0, 0, 0)$

2. Express the following as linear combinations of $\mathbf{u} = (2, 1, 4)$, $\mathbf{v} = (1, -1, 3)$, and $\mathbf{w} = (3, 2, 5)$.

   a. $(-9, -7, -15)$ **b.** $(6, 11, 6)$ **c.** $(0, 0, 0)$

3. Which of the following are linear combinations of

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

   a. $\begin{bmatrix} 6 & -8 \\ -1 & -8 \end{bmatrix}$ **b.** $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ **c.** $\begin{bmatrix} -1 & 5 \\ 7 & 1 \end{bmatrix}$

4. In each part, determine whether the polynomial is a linear combination of

   $$\mathbf{p}_1 = 2 + x + x^2, \quad \mathbf{p}_2 = 1 - x^2, \quad \mathbf{p}_3 = 1 + 2x.$$

   a. $1 + x$ **b.** $1 + x^2$ **c.** $1 + x + x^2$

5. In each part, express the vector as a linear combination of

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

   a. $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ **b.** $\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$

6. In each part express the vector as a linear combination of $\mathbf{p}_1 = 2 + x + 4x^2$, $\mathbf{p}_2 = 1 - x + 3x^2$, and $\mathbf{p}_3 = 3 + 2x + 5x^2$.

   a. $-9 - 7x - 15x^2$ **b.** $6 + 11x + 6x^2$
   c. 0 **d.** $7 + 8x + 9x^2$

7. In each part, determine whether the vectors span R^3.

   a. $\mathbf{v}_1 = (2, 2, 2)$, $\mathbf{v}_2 = (0, 0, 3)$, $\mathbf{v}_3 = (0, 1, 1)$
   b. $\mathbf{v}_1 = (2, -1, 3)$, $\mathbf{v}_2 = (4, 1, 2)$, $\mathbf{v}_3 = (8, -1, 8)$

8. Suppose that $\mathbf{v}_1 = (2, 1, 0, 3)$, $\mathbf{v}_2 = (3, -1, 5, 2)$, and $\mathbf{v}_3 = (-1, 0, 2, 1)$. Which of the following vectors are in $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$?

   a. $(2, 3, -7, 3)$ **b.** $(0, 0, 0, 0)$
   c. $(1, 1, 1, 1)$ **d.** $(-4, 6, -13, 4)$

9. Determine whether the following polynomials span P_2.

   $$\mathbf{p}_1 = 1 - x + 2x^2, \quad \mathbf{p}_2 = 3 + x,$$
   $$\mathbf{p}_3 = 5 - x + 4x^2, \quad \mathbf{p}_4 = -2 - 2x + 2x^2$$

10. Determine whether the following polynomials span P_2.

   $$\mathbf{p}_1 = 1 + x, \quad \mathbf{p}_2 = 1 - x,$$
   $$\mathbf{p}_3 = 1 + x + x^2, \quad \mathbf{p}_4 = 2 - x^2$$

11. In each part, determine whether the matrices span M_{22}.

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

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

c. $\begin{bmatrix}1&0\\0&0\end{bmatrix}, \begin{bmatrix}1&1\\0&0\end{bmatrix}, \begin{bmatrix}1&1\\1&0\end{bmatrix}, \begin{bmatrix}1&1\\1&1\end{bmatrix}$

12. Let $T_A : R^2 \to R^2$ be multiplication by A. Determine whether the vector $\mathbf{u} = (1, 2)$ is in the span of $\{T_A(\mathbf{e}_1), T_A(\mathbf{e}_2)\}$.

a. $A = \begin{bmatrix}1&2\\0&1\end{bmatrix}$ **b.** $A = \begin{bmatrix}1&1\\1&1\end{bmatrix}$

13. Let $T_A : R^2 \to R^3$ be multiplication by A. Determine whether the vector $\mathbf{u} = (1, 1, 1)$ is in the span of $\{T_A(\mathbf{e}_1), T_A(\mathbf{e}_2)\}$.

a. $A = \begin{bmatrix}0&2\\1&-2\\1&0\end{bmatrix}$ **b.** $A = \begin{bmatrix}0&2\\1&1\\2&0\end{bmatrix}$

14. Let $\mathbf{f} = \cos^2 x$ and $\mathbf{g} = \sin^2 x$. Which of the following lie in the space spanned by $\mathbf{f}$ and $\mathbf{g}$?

a. $\cos 2x$ **b.** $3 + x^2$ **c.** 1 **d.** $\sin x$ **e.** 0

15. Let W be the solution space to the system $A\mathbf{x} = \mathbf{0}$. Determine whether the set $\{\mathbf{u}, \mathbf{v}\}$ spans W.

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

a. $\mathbf{u} = (1, 0, -1, 0)$, $\mathbf{v} = (0, 1, 0, -1)$

b. $\mathbf{u} = (1, 0, -1, 0)$, $\mathbf{v} = (1, 1, -1, -1)$

16. Let W be the solution space to the system $A\mathbf{x} = \mathbf{0}$. Determine whether the set $\{\mathbf{u}, \mathbf{v}\}$ spans W.

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

a. $\mathbf{u} = (1, 1, 1, 0)$, $\mathbf{v} = (0, -1, 0, 1)$

b. $\mathbf{u} = (0, 1, 1, 0)$, $\mathbf{v} = (1, 0, 1, 1)$

17. In each part, let $T_A : R^2 \to R^2$ be multiplication by A, and let $\mathbf{u}_1 = (1, 2)$ and $\mathbf{u}_2 = (-1, 1)$. Determine whether the set $\{T_A(\mathbf{u}_1), T_A(\mathbf{u}_2)\}$ spans R^2.

a. $A = \begin{bmatrix}1&-1\\0&2\end{bmatrix}$ **b.** $A = \begin{bmatrix}1&-1\\-2&2\end{bmatrix}$

18. In each part, let $T_A : R^3 \to R^2$ be multiplication by A, and let $\mathbf{u}_1 = (0, 1, 1)$ and $\mathbf{u}_2 = (2, -1, 1)$ and $\mathbf{u}_3 = (1, 1, -2)$. Determine whether the set $\{T_A(\mathbf{u}_1), T_A(\mathbf{u}_2), T_A(\mathbf{u}_3)\}$ spans R^2.

a. $A = \begin{bmatrix}1&1&0\\0&1&-1\end{bmatrix}$ **b.** $A = \begin{bmatrix}0&1&0\\1&1&-3\end{bmatrix}$

19. Let $\mathbf{p}_1 = 1 + x^2$, $\mathbf{p}_2 = 1 + x + x^2$, and $\mathbf{q}_1 = 2x$, $\mathbf{q}_2 = 1 + x^2$. Use Theorem 4.3.2 to show that $\text{span}\{\mathbf{p}_1, \mathbf{p}_2\} = \text{span}\{\mathbf{q}_1, \mathbf{q}_2\}$.

20. Let $\mathbf{v}_1 = (1, 6, 4)$, $\mathbf{v}_2 = (2, 4, -1)$, $\mathbf{v}_3 = (-1, 2, 5)$, and $\mathbf{w}_1 = (1, -2, -5)$, $\mathbf{w}_2 = (0, 8, 9)$. Use Theorem 4.3.2 to show that $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \text{span}\{\mathbf{w}_1, \mathbf{w}_2\}$.

21. Let V and W be subspaces of R^2 that are spanned by $(3, 1)$ and $(2, 1)$, respectively. Find a vector $\mathbf{v}$ in V and a vector $\mathbf{w}$ in W for which $\mathbf{v} + \mathbf{w} = (3, 5)$.

22. Let V be the solution space of the equation $4x - y + 2z = 0$, and let W be the subspace of R^3 spanned by $(1, 1, 1)$. Find a vector $\mathbf{v}$ in V and a vector $\mathbf{w}$ in W for which

$$\mathbf{v} + \mathbf{w} = (1, 0, 1)$$

Working with Proofs

23. Prove that if $\{\mathbf{u}, \mathbf{v}\}$ spans the vector space V, then $\{\mathbf{u}, \mathbf{u} + \mathbf{v}\}$ spans V.

24. Prove Theorem 4.3.2.

True-False Exercises

TF. In parts (a)–(g) determine whether the statement is true or false, and justify your answer.

a. An expression of the form $k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots k_r\mathbf{v}_r$ is called a linear combination.

b. The span of a single vector in R^2 is a line.

c. The span of two vectors in R^3 is a plane.

d. The span of a nonempty set S of vectors in V is the smallest subspace of V that contains S.

e. The span of any finite set of vectors in a vector space is closed under addition and scalar multiplication.

f. Two subsets of a vector space V that span the same subspace of V must be equal.

g. The polynomials $x - 1$, $(x - 1)^2$, and $(x - 1)^3$ span P_3.

Working with Technology

T1. Recall from Theorem 1.3.1 that a product $A\mathbf{x}$ can be expressed as a linear combination of the column vectors of the matrix A in which the coefficients are the entries of $\mathbf{x}$. Use matrix multiplication to compute

$$\mathbf{v} = 6(8, -2, 1, -4) + 17(-3, 9, 11, 6) - 9(13, -1, 2, 4)$$

T2. Use the idea in Exercise T1 and matrix multiplication to determine whether the polynomial

$$\mathbf{p} = 1 + x + x^2 + x^3$$

is in the span of

$$\mathbf{p}_1 = 8 - 2x + x^2 - 4x^3, \quad \mathbf{p}_2 = -3 + 9x + 11x^2 + 6x^3,$$
$$\mathbf{p}_3 = 13 - x + 2x^2 + 4x^3$$

T3. For the vectors that follow, determine whether

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \text{span}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$$

$$\mathbf{v}_1 = (-1, 2, 0, 1, 3), \quad \mathbf{v}_2 = (7, 4, 6, -3, 1),$$
$$\mathbf{v}_3 = (-5, 3, 1, 2, 4)$$
$$\mathbf{w}_1 = (-6, 5, 1, 3, 7), \quad \mathbf{w}_2 = (6, 6, 6, -2, 4),$$
$$\mathbf{w}_3 = (2, 7, 7, -1, 5)$$

4.4 Linear Independence

In this section we will consider the question of whether the vectors in a given set are interrelated in the sense that one or more of them can be expressed as a linear combination of the others. This is important to know in applications because the existence of such relationships often signals that some kind of complication is likely to occur.

Linear Independence and Dependence

In a rectangular xy-coordinate system every vector in the plane can be expressed in exactly one way as a linear combination of the standard unit vectors. For example, the only way to express the vector $(3, 2)$ as a linear combination of $\mathbf{i} = (1, 0)$ and $\mathbf{j} = (0, 1)$ is

$$(3, 2) = 3(1, 0) + 2(0, 1) = 3\mathbf{i} + 2\mathbf{j} \tag{1}$$

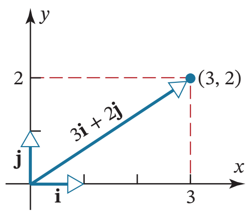


FIGURE 4.4.1

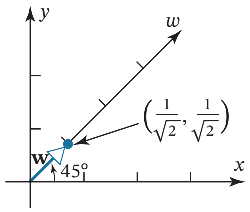


FIGURE 4.4.2

(**Figure 4.4.1**). Suppose, however, that we were to introduce a third coordinate axis that makes an angle of 45° with the x-axis. Call it the w-axis. As illustrated in **Figure 4.4.2**, the unit vector along the w-axis is

$$\mathbf{w} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

Whereas Formula (1) shows the only way to express the vector $(3, 2)$ as a linear combination of $\mathbf{i}$ and $\mathbf{j}$, there are infinitely many ways to express this vector as a linear combination of $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{w}$. Three possibilities are

$$(3, 2) = 3(1, 0) + 2(0, 1) + 0\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = 3\mathbf{i} + 2\mathbf{j} + 0\mathbf{w}$$

$$(3, 2) = 2(1, 0) + (0, 1) + \sqrt{2}\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = 2\mathbf{i} + \mathbf{j} + \sqrt{2}\mathbf{w}$$

$$(3, 2) = 4(1, 0) + 3(0, 1) - \sqrt{2}\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = 4\mathbf{i} + 3\mathbf{j} - \sqrt{2}\mathbf{w}$$

In short, by introducing a superfluous axis we created the complication of having multiple ways of assigning coordinates to points in the plane. What makes the vector $\mathbf{w}$ superfluous is the fact that it can be expressed as a linear combination of the vectors $\mathbf{i}$ and $\mathbf{j}$, namely,

$$\mathbf{w} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}}\mathbf{i} + \frac{1}{\sqrt{2}}\mathbf{j}$$

This leads to the following definition.

Definition 1

If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a set of two or more vectors in a vector space V, then S is said to be a ***linearly independent set*** if no vector in S can be expressed as a linear combination of the others. A set that is not linearly independent is said to be ***linearly dependent***. If S has only one vector, we will agree that it is linearly independent if and only if that vector is nonzero.

In the case where the set S in Definition 1 has only one vector, we will agree that S is linearly independent if and only if that vector is nonzero.

In general, the most efficient way to determine whether a set is linearly independent or not is to use the following theorem whose proof is given at the end of this section.

Theorem 4.4.1

A nonempty set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ in a vector space V is linearly independent if and only if the only coefficients satisfying the vector equation

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_r\mathbf{v}_r = \mathbf{0}$$

are $k_1 = 0, k_2 = 0, \ldots, k_r = 0$.

EXAMPLE 1 | Linear Independence of the Standard Unit Vectors in R^n

The most basic linearly independent set in R^n is the set of standard unit vectors

$$\mathbf{e}_1 = (1, 0, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, 0, \ldots, 0), \ldots, \quad \mathbf{e}_n = (0, 0, 0, \ldots, 1)$$

To illustrate this in R^3, consider the standard unit vectors

$$\mathbf{i} = (1, 0, 0), \quad \mathbf{j} = (0, 1, 0), \quad \mathbf{k} = (0, 0, 1)$$

To prove linear independence we must show that the only coefficients satisfying the vector equation

$$k_1\mathbf{i} + k_2\mathbf{j} + k_3\mathbf{k} = \mathbf{0}$$

are $k_1 = 0, k_2 = 0, k_3 = 0$. But this becomes evident by writing this equation in its component form

$$(k_1, k_2, k_3) = (0, 0, 0)$$

You should have no trouble adapting this argument to establish the linear independence of the standard unit vectors in R^n.

EXAMPLE 2 | Linear Independence in R^3

Determine whether the vectors

$$\mathbf{v}_1 = (1, -2, 3), \quad \mathbf{v}_2 = (5, 6, -1), \quad \mathbf{v}_3 = (3, 2, 1) \tag{2}$$

are linearly independent or linearly dependent in R^3.

Solution The linear independence or dependence of these vectors is determined by whether the vector equation

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + k_3\mathbf{v}_3 = \mathbf{0} \tag{3}$$

can be satisfied with coefficients that are not all zero. To see whether this is so, let us rewrite (3) in the component form

$$k_1(1, -2, 3) + k_2(5, 6, -1) + k_3(3, 2, 1) = (0, 0, 0)$$

Equating corresponding components on the two sides yields the homogeneous linear system

$$\begin{aligned} k_1 + 5k_2 + 3k_3 &= 0 \\ -2k_1 + 6k_2 + 2k_3 &= 0 \\ 3k_1 - k_2 + k_3 &= 0 \end{aligned} \tag{4}$$

Thus, our problem reduces to determining whether this system has nontrivial solutions. There are various ways to do this; one possibility is to simply solve the system, which yields

$$k_1 = -\tfrac{1}{2}t, \quad k_2 = -\tfrac{1}{2}t, \quad k_3 = t$$

(we omit the details). This shows that the system has nontrivial solutions and hence that the vectors are linearly dependent. A second method for establishing the linear dependence is to take advantage of the fact that the coefficient matrix

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

is square and compute its determinant. We leave it for you to show that $\det(A) = 0$ from which it follows that (4) has nontrivial solutions by parts (*b*) and (*g*) of Theorem 2.3.8.

Because we have established that the vectors $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$ in (2) are linearly dependent, we know that at least one of them is a linear combination of the others. We leave it for you to confirm, for example, that

$$\mathbf{v}_3 = \tfrac{1}{2}\mathbf{v}_1 + \tfrac{1}{2}\mathbf{v}_2$$

EXAMPLE 3 | Linear Independence in R^4

Determine whether the vectors

$$\mathbf{v}_1 = (1, 2, 2, -1), \quad \mathbf{v}_2 = (4, 9, 9, -4), \quad \mathbf{v}_3 = (5, 8, 9, -5)$$

in R^4 are linearly dependent or linearly independent.

Solution The linear independence or linear dependence of these vectors is determined by whether there exist nontrivial solutions of the vector equation

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + k_3\mathbf{v}_3 = \mathbf{0}$$

or, equivalently, of

$$k_1(1, 2, 2, -1) + k_2(4, 9, 9, -4) + k_3(5, 8, 9, -5) = (0, 0, 0, 0)$$

Equating corresponding components on the two sides yields the homogeneous linear system

$$\begin{aligned} k_1 + 4k_2 + 5k_3 &= 0 \\ 2k_1 + 9k_2 + 8k_3 &= 0 \\ 2k_1 + 9k_2 + 9k_3 &= 0 \\ -k_1 - 4k_2 - 5k_3 &= 0 \end{aligned}$$

We leave it for you to show that this system has only the trivial solution

$$k_1 = 0, \quad k_2 = 0, \quad k_3 = 0$$

from which you can conclude that $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$ are linearly independent.

EXAMPLE 4 | An Important Linearly Independent Set in P_n

Show that the polynomials

$$1, \quad x, \quad x^2, \ldots, \quad x^n$$

form a linearly independent set in P_n.

Solution For convenience, let us denote the polynomials as

$$\mathbf{p}_0 = 1, \quad \mathbf{p}_1 = x, \quad \mathbf{p}_2 = x^2, \ldots, \quad \mathbf{p}_n = x^n$$

We must show that the only coefficients satisfying the vector equation

$$a_0\mathbf{p}_0 + a_1\mathbf{p}_1 + a_2\mathbf{p}_2 + \cdots + a_n\mathbf{p}_n = \mathbf{0} \tag{5}$$

are

$$a_0 = a_1 = a_2 = \cdots = a_n = 0$$

But (5) is equivalent to the statement that

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n = 0 \tag{6}$$

for all x in $(-\infty, \infty)$, so we must show that this is true if and only if each coefficient in (6) is zero. To see that this is so, recall from algebra that a nonzero polynomial of degree n has at most n distinct roots. That being the case, each coefficient in (6) must be zero, for otherwise the left side of the equation would be a nonzero polynomial with infinitely many roots. Thus, (5) has only the trivial solution.

The following example shows that the problem of determining whether a given set of vectors in P_n is linearly independent or linearly dependent can be reduced to determining whether a certain set of vectors in R^n is linearly dependent or independent.

EXAMPLE 5 | Linear Independence of Polynomials

Determine whether the polynomials

$$\mathbf{p}_1 = 1 - x, \quad \mathbf{p}_2 = 5 + 3x - 2x^2, \quad \mathbf{p}_3 = 1 + 3x - x^2$$

are linearly dependent or linearly independent in P_2.

Solution The linear independence or dependence of these vectors is determined by whether the vector equation

$$k_1\mathbf{p}_1 + k_2\mathbf{p}_2 + k_3\mathbf{p}_3 = \mathbf{0} \tag{7}$$

can be satisfied with coefficients that are not all zero. To see whether this is so, let us rewrite (7) in its polynomial form

$$k_1(1 - x) + k_2(5 + 3x - 2x^2) + k_3(1 + 3x - x^2) = 0 \tag{8}$$

or, equivalently, as

$$(k_1 + 5k_2 + k_3) + (-k_1 + 3k_2 + 3k_3)x + (-2k_2 - k_3)x^2 = 0$$

Since this equation must be satisfied by all x in $(-\infty, \infty)$, each coefficient must be zero (as explained in the previous example). Thus, the linear dependence or independence of the given polynomials hinges on whether the following linear system has a nontrivial solution:

$$\begin{aligned} k_1 + 5k_2 + k_3 &= 0 \\ -k_1 + 3k_2 + 3k_3 &= 0 \\ - 2k_2 - k_3 &= 0 \end{aligned} \tag{9}$$

We leave it for you to show that this linear system has nontrivial solutions either by solving it directly or by showing that the coefficient matrix has determinant zero. Thus, the set $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ is linearly dependent.

In Example 5, what relationship do you see between the coefficients of the given polynomials and the column vectors of the coefficient matrix of system (9)?

The following useful theorem is concerned with the linear independence of sets with two vectors and sets that contain the zero vector.

Theorem 4.4.2

(*a*) A set with finitely many vectors that contains $\mathbf{0}$ is linearly dependent.

(*b*) A set with exactly two vectors is linearly independent if and only if neither vector is a scalar multiple of the other.

We will prove part (a) and leave part (b) as an exercise.

Proof **(a)** For any vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$, the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r, \mathbf{0}\}$ is linearly dependent since the equation

$$0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_r + 1(\mathbf{0}) = \mathbf{0}$$

expresses $\mathbf{0}$ as a linear combination of the vectors in S with coefficients that are not all zero. ■

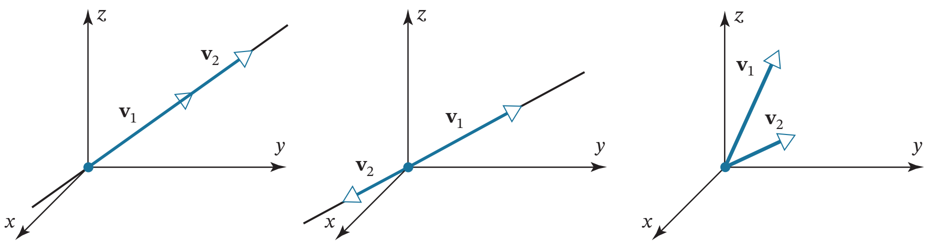
EXAMPLE 6 | Linear Independence of Two Functions

The functions $\mathbf{f}_1 = x$ and $\mathbf{f}_2 = \sin x$ are linearly independent vectors in $F(-\infty, \infty)$ since neither function is a scalar multiple of the other. On the other hand, the two functions $\mathbf{g}_1 = \sin 2x$ and $\mathbf{g}_2 = \sin x \cos x$ are linearly dependent because the trigonometric identity $\sin 2x = 2 \sin x \cos x$ reveals that $\mathbf{g}_1$ and $\mathbf{g}_2$ are scalar multiples of each other.

A Geometric Interpretation of Linear Independence

Linear independence has the following useful geometric interpretations in R^2 and R^3:

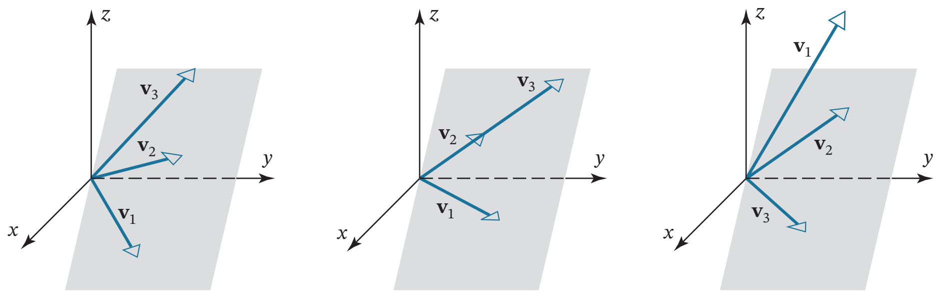
- Two vectors in R^2 or R^3 are linearly independent if and only if they do not lie on the same line when they have their initial points at the origin. Otherwise one would be a scalar multiple of the other (**Figure 4.4.3**).



(a) Linearly dependent **(b)** Linearly dependent **(c)** Linearly independent

FIGURE 4.4.3

- Three vectors in R^3 are linearly independent if and only if they do not lie in the same plane when they have their initial points at the origin. Otherwise at least one would be a linear combination of the other two (**Figure 4.4.4**).



(a) Linearly dependent **(b)** Linearly dependent **(c)** Linearly independent

FIGURE 4.4.4

At the beginning of this section we observed that a third coordinate axis in R^2 is superfluous by showing that a unit vector along such an axis would have to be expressible as a linear combination of unit vectors along the positive x- and y-axis. That result is a consequence of the next theorem, which shows that there can be at most n vectors in any linearly independent set R^n.

Theorem 4.4.3

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a set of vectors in R^n. If $r > n$, then S is linearly dependent.

Proof Suppose that

$$\begin{aligned} \mathbf{v}_1 &= (v_{11}, v_{12}, \ldots, v_{1n}) \\ \mathbf{v}_2 &= (v_{21}, v_{22}, \ldots, v_{2n}) \\ &\;\;\vdots \qquad\quad \vdots \\ \mathbf{v}_r &= (v_{r1}, v_{r2}, \ldots, v_{rn}) \end{aligned}$$

and consider the equation

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_r\mathbf{v}_r = \mathbf{0}$$

If we express both sides of this equation in terms of components and then equate the corresponding components, we obtain the system

$$\begin{array}{ccccccc} v_{11}k_1 & + & v_{21}k_2 & + \cdots + & v_{r1}k_r & = & 0 \\ v_{12}k_1 & + & v_{22}k_2 & + \cdots + & v_{r2}k_r & = & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ v_{1n}k_1 & + & v_{2n}k_2 & + \cdots + & v_{rn}k_r & = & 0 \end{array}$$

This is a homogeneous system of n equations in the r unknowns $k_1, \ldots, k_r$. Since $r > n$, Theorem 1.2.2 implies that the system has nontrivial solutions, so $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a linearly dependent set. ■

It follows from Theorem 4.4.3 that a set in R^2 with more than two vectors is linearly dependent and a set in R^3 with more than three vectors is linearly dependent.

EXAMPLE 7 | Linear Independence of Row Vectors in a Row Echelon Form

It is an important fact that the nonzero row vectors of a matrix in row echelon or reduced row echelon form are linearly independent. To suggest how a general proof might go, consider the matrix

$$R = \begin{bmatrix} 1 & a_{12} & a_{13} & a_{14} \\ 0 & 1 & a_{23} & a_{24} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which is in row echelon form for all choices of the a's. Denoting the row vectors by $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$, we must show that the only solution of the vector equation

$$c_1\mathbf{r}_1 + c_2\mathbf{r}_2 + c_3\mathbf{r}_3 = \mathbf{0} \tag{10}$$

is the trivial solution $c_1 = c_2 = c_3 = 0$. We can do this by writing (10) in the row-vector form

$$\begin{bmatrix} c_1 & c_1a_{12} + c_2 & c_1a_{13} + c_2a_{23} & c_1a_{14} + c_2a_{24} + c_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$$

and comparing corresponding components. We see from the first component that $c_1 = 0$, and from the second component that $c_2 = 0$, and hence from the fourth component that $c_3 = 0$. Thus, (10) has only the trivial solution.

Linear Independence of Functions

CALCULUS REQUIRED

Sometimes linear dependence of functions can be deduced from known identities. For example, the functions

$$\mathbf{f}_1 = \sin^2 x, \quad \mathbf{f}_2 = \cos^2 x, \quad \text{and} \quad \mathbf{f}_3 = 5$$

form a linearly dependent set in $F(-\infty, \infty)$, since the equation

$$\begin{aligned} 5\mathbf{f}_1 + 5\mathbf{f}_2 - \mathbf{f}_3 &= 5\sin^2 x + 5\cos^2 x - 5 \\ &= 5(\sin^2 x + \cos^2 x) - 5 = \mathbf{0} \end{aligned}$$

expresses the vector $\mathbf{0}$ as a linear combination of $\mathbf{f}_1$, $\mathbf{f}_2$, and $\mathbf{f}_3$ with coefficients that are not all zero.

However, it is relatively rare that linear independence or dependence of functions can be ascertained by algebraic or trigonometric methods. To make matters worse, there is no general method for doing that either. That said, there does exist a theorem that can be useful in certain cases. The following definition is needed for that theorem.

Definition 2

If $\mathbf{f}_1 = f_1(x), \mathbf{f}_2 = f_2(x), \ldots, \mathbf{f}_n = f_n(x)$ are functions that are $n-1$ times differentiable on the interval $(-\infty, \infty)$, then the determinant

$$W(x) = \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ \vdots & \vdots & & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{vmatrix}$$

is called the ***Wronskian*** of $f_1, f_2, \ldots, f_n$.

Suppose for the moment that $\mathbf{f}_1 = f_1(x), \mathbf{f}_2 = f_2(x), \ldots, \mathbf{f}_n = f_n(x)$ are *linearly dependent* vectors in $C^{(n-1)}(-\infty, \infty)$. This implies that the vector equation

$$k_1\mathbf{f}_1 + k_2\mathbf{f}_2 + \cdots + k_n\mathbf{f}_n = \mathbf{0}$$

is satisfied by values of the coefficients $k_1, k_2, \ldots, k_n$ that are not all zero, and for these coefficients the equation

$$k_1 f_1(x) + k_2 f_2(x) + \cdots + k_n f_n(x) = 0$$

is satisfied for all x in $(-\infty, \infty)$. Using this equation together with those that result by differentiating it $n-1$ times we obtain the linear system

$$\begin{array}{lllll} k_1 f_1(x) & + k_2 f_2(x) & + \cdots + k_n f_n(x) & = 0 \\ k_1 f_1'(x) & + k_2 f_2'(x) & + \cdots + k_n f_n'(x) & = 0 \\ \vdots & \vdots & \vdots & \vdots \\ k_1 f_1^{(n-1)}(x) & + k_2 f_2^{(n-1)}(x) & + \cdots + k_n f_n^{(n-1)}(x) & = 0 \end{array}$$

Thus, the assumed linear dependence of $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$ implies that the linear system

$$\begin{bmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ \vdots & \vdots & & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{11}$$

has a nontrivial solution for every x in the interval $(-\infty, \infty)$, and this in turn implies that the determinant of the coefficient matrix of (11) is zero for every such x. Thus, the assumed linear independence of $f_1, f_2, \ldots, f_n$ implies that the Wronskian of these functions is identically zero on $(-\infty, \infty)$; or stated in contrapositive form (see Appendix A), if the Wronskian is not identically zero on $(-\infty, \infty)$, then the functions must be linearly dependent. Thus, we have the following result.

Theorem 4.4.4

If the functions $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$ have $n-1$ continuous derivatives on the interval $(-\infty, \infty)$, and if the Wronskian of these functions is not identically zero on $(-\infty, \infty)$, then these functions form a linearly independent set of vectors in $C^{(n-1)}(-\infty, \infty)$.

Warning The converse of Theorem 4.4.4 is false. If the Wronskian of $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$ is identically zero on $(-\infty, \infty)$, then no conclusion can be reached about the linear independence of $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$—this set of vectors may be linearly independent or linearly dependent.

In Example 6 we showed that x and $\sin x$ are linearly independent functions by observing that neither is a scalar multiple of the other. The following example shows that this is consistent with Theorem 4.4.4.

EXAMPLE 8 | Linear Independence Using the Wronskian

Use the Wronskian to show that $\mathbf{f}_1 = x$ and $\mathbf{f}_2 = \sin x$ are linearly independent vectors in $C^\infty(-\infty, \infty)$.

Solution The Wronskian is

$$W(x) = \begin{vmatrix} x & \sin x \\ 1 & \cos x \end{vmatrix} = x\cos x - \sin x$$

This function is not identically zero on the interval $(-\infty, \infty)$ since, for example,

$$W\left(\frac{\pi}{2}\right) = \frac{\pi}{2}\cos\left(\frac{\pi}{2}\right) - \sin\left(\frac{\pi}{2}\right) = -1$$

Thus, the functions are linearly independent.

Historical Note

Józef Hoëné de Wroński (1778–1853)

The Polish-French mathematician Józef Hoëné de Wroński was born Józef Hoëné and adopted the name Wroński after he married. Wroński's life was fraught with controversy and conflict, which some say was due to psychopathic tendencies and his exaggeration of the importance of his own work. Although Wroński's work was dismissed as rubbish for many years, and much of it was indeed erroneous, some of his ideas contained hidden brilliance and have survived. In addition to his purely mathematical work, he designed a caterpillar vehicle to compete with trains (though it was never manufactured) and did research on the famous problem of determining the longitude of a ship at sea. His final years were spent in poverty.

[Image: © TopFoto/The Image Works]

EXAMPLE 9 | Linear Independence Using the Wronskian

Use the Wronskian to show that $\mathbf{f}_1 = 1, \mathbf{f}_2 = e^x$, and $\mathbf{f}_3 = e^{2x}$ are linearly independent vectors in $C^\infty(-\infty, \infty)$.

Solution The Wronskian is

$$W(x) = \begin{vmatrix} 1 & e^x & e^{2x} \\ 0 & e^x & 2e^{2x} \\ 0 & e^x & 4e^{2x} \end{vmatrix} = 2e^{3x}$$

This function is obviously not identically zero on $(-\infty, \infty)$, so $\mathbf{f}_1$, $\mathbf{f}_2$, and $\mathbf{f}_3$ form a linearly independent set.

OPTIONAL: We will close this section by proving Theorem 4.4.1.

Proof of Theorem 4.4.1 We will prove this theorem in the case where the set S has two or more vectors, and leave the case where S has only one vector as an exercise. Assume first that S is linearly independent. We will show that if the equation

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_r\mathbf{v}_r = \mathbf{0} \tag{12}$$

can be satisfied with coefficients that are not all zero, then at least one of the vectors in S must be expressible as a linear combination of the others, thereby contradicting the assumption of linear independence. To be specific, suppose that $k_1 \neq 0$. Then we can rewrite (12) as

$$\mathbf{v}_1 = \left(-\frac{k_2}{k_1}\right)\mathbf{v}_2 + \cdots + \left(-\frac{k_r}{k_1}\right)\mathbf{v}_r$$

which expresses $\mathbf{v}_1$ as a linear combination of the other vectors in S.

Conversely, we must show that if the only coefficients satisfying (12) are

$$k_1 = 0, \quad k_2 = 0, \ldots, \quad k_r = 0$$

then the vectors in S must be linearly independent. But if this were true of the coefficients and the vectors were not linearly independent, then at least one of them would be expressible as a linear combination of the others, say

$$\mathbf{v}_1 = c_2\mathbf{v}_2 + \cdots + c_r\mathbf{v}_r$$

which we can rewrite as

$$\mathbf{v}_1 + (-c_2)\mathbf{v}_2 + \cdots + (-c_r)\mathbf{v}_r = \mathbf{0}$$

But this contradicts our assumption that (12) can only be satisfied by coefficients that are all zero. Thus, the vectors in S must be linearly independent. ■

Exercise Set 4.4

1. Explain why the following form linearly dependent sets of vectors. (Solve this problem by inspection.)

   a. $\mathbf{u}_1 = (-1, 2, 4)$ and $\mathbf{u}_2 = (5, -10, -20)$ in R^3

   b. $\mathbf{u}_1 = (3, -1)$, $\mathbf{u}_2 = (4, 5)$, $\mathbf{u}_3 = (-4, 7)$ in R^2

   c. $\mathbf{p}_1 = 3 - 2x + x^2$ and $\mathbf{p}_2 = 6 - 4x + 2x^2$ in P_2

   d. $A = \begin{bmatrix} -3 & 4 \\ 2 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -4 \\ -2 & 0 \end{bmatrix}$ in M_{22}

2. In each part, determine whether the vectors are linearly independent or are linearly dependent in R^3.

   a. $(-3, 0, 4)$, $(5, -1, 2)$, $(1, 1, 3)$

   b. $(-2, 0, 1)$, $(3, 2, 5)$, $(6, -1, 1)$, $(7, 0, -2)$

3. In each part, determine whether the vectors are linearly independent or are linearly dependent in R^4.

   a. $(3, 8, 7, -3)$, $(1, 5, 3, -1)$, $(2, -1, 2, 6)$, $(4, 2, 6, 4)$

   b. $(3, 0, -3, 6)$, $(0, 2, 3, 1)$, $(0, -2, -2, 0)$, $(-2, 1, 2, 1)$

4. In each part, determine whether the vectors are linearly independent or are linearly dependent in P_2.

a. $2 - x + 4x^2,\ 3 + 6x + 2x^2,\ 2 + 10x - 4x^2$

b. $1 + 3x + 3x^2,\ x + 4x^2,\ 5 + 6x + 3x^2,\ 7 + 2x - x^2$

5. In each part, determine whether the matrices are linearly independent or dependent.

a. $\begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}$ in M_{22}

b. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ in M_{23}

6. Determine all values of k for which the following matrices are linearly independent in M_{22}.

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

7. In each part, determine whether the three vectors lie in a plane in R^3.

a. $\mathbf{v}_1 = (2, -2, 0),\ \mathbf{v}_2 = (6, 1, 4),\ \mathbf{v}_3 = (2, 0, -4)$

b. $\mathbf{v}_1 = (-6, 7, 2),\ \mathbf{v}_2 = (3, 2, 4),\ \mathbf{v}_3 = (4, -1, 2)$

8. In each part, determine whether the three vectors lie on the same line in R^3.

a. $\mathbf{v}_1 = (-1, 2, 3),\ \mathbf{v}_2 = (2, -4, -6),\ \mathbf{v}_3 = (-3, 6, 0)$

b. $\mathbf{v}_1 = (2, -1, 4),\ \mathbf{v}_2 = (4, 2, 3),\ \mathbf{v}_3 = (2, 7, -6)$

c. $\mathbf{v}_1 = (4, 6, 8),\ \mathbf{v}_2 = (2, 3, 4),\ \mathbf{v}_3 = (-2, -3, -4)$

9. a. Show that the three vectors $\mathbf{v}_1 = (0, 3, 1, -1)$, $\mathbf{v}_2 = (6, 0, 5, 1)$, and $\mathbf{v}_3 = (4, -7, 1, 3)$ form a linearly dependent set in R^4.

b. Express each vector in part (a) as a linear combination of the other two.

10. a. Show that the vectors $\mathbf{v}_1 = (1, 2, 3, 4)$, $\mathbf{v}_2 = (0, 1, 0, -1)$, and $\mathbf{v}_3 = (1, 3, 3, 3)$ form a linearly dependent set in R^4.

b. Express each vector in part (a) as a linear combination of the other two.

11. For which real values of λ do the following vectors form a linearly dependent set in R^3?

$$\mathbf{v}_1 = \left(\lambda, -\tfrac{1}{2}, -\tfrac{1}{2}\right),\ \mathbf{v}_2 = \left(-\tfrac{1}{2}, \lambda, -\tfrac{1}{2}\right),\ \mathbf{v}_3 = \left(-\tfrac{1}{2}, -\tfrac{1}{2}, \lambda\right)$$

12. Under what conditions is a set with one vector linearly independent?

13. In each part, let $T_A: R^2 \to R^2$ be multiplication by A, and let $\mathbf{u}_1 = (1, 2)$ and $\mathbf{u}_2 = (-1, 1)$. Determine whether the set $\{T_A(\mathbf{u}_1), T_A(\mathbf{u}_2)\}$ is linearly independent in R^2.

a. $A = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \end{bmatrix}$

14. In each part, let $T_A: R^3 \to R^3$ be multiplication by A, and let $\mathbf{u}_1 = (1, 0, 0)$, $\mathbf{u}_2 = (2, -1, 1)$, and $\mathbf{u}_3 = (0, 1, 1)$. Determine whether the set $\{T_A(\mathbf{u}_1), T_A(\mathbf{u}_2), T_A(\mathbf{u}_3)\}$ is linearly independent in R^3.

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

15. Are the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ in part (*a*) of the accompanying figure linearly independent? What about those in part (*b*)? Explain.

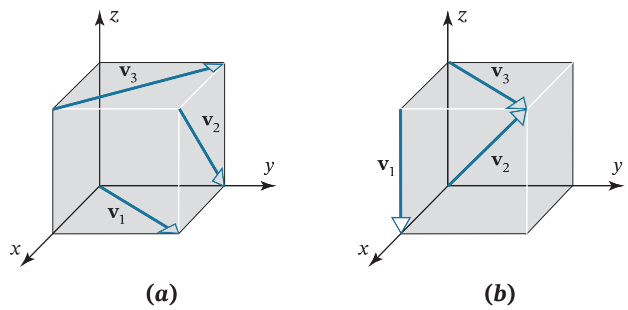


FIGURE Ex-15

16. By using appropriate identities, where required, determine which of the following sets of vectors in $F(-\infty, \infty)$ are linearly dependent.

a. $6,\ 3\sin^2 x,\ 2\cos^2 x$ **b.** $x,\ \cos x$

c. $1,\ \sin x,\ \sin 2x$ **d.** $\cos 2x,\ \sin^2 x,\ \cos^2 x$

e. $(3 - x)^2,\ x^2 - 6x,\ 5$ **f.** $0,\ \cos^3 \pi x,\ \sin^5 3\pi x$

17. (*Calculus required*) The functions

$$f_1(x) = x \quad \text{and} \quad f_2(x) = \cos x$$

are linearly independent in $F(-\infty, \infty)$ because neither function is a scalar multiple of the other. Confirm the linear independence using the Wronskian.

18. (*Calculus required*) The functions

$$f_1(x) = \sin x \quad \text{and} \quad f_2(x) = \cos x$$

are linearly independent in $F(-\infty, \infty)$ because neither function is a scalar multiple of the other. Confirm the linear independence using the Wronskian.

19. (*Calculus required*) Use the Wronskian to show that the following sets of vectors are linearly independent.

a. $1,\ x,\ e^x$ **b.** $1,\ x,\ x^2$

20. (*Calculus required*) Use the Wronskian to show that the functions $f_1(x) = e^x$, $f_2(x) = xe^x$, and $f_3(x) = x^2e^x$ are linearly independent vectors in $C^\infty(-\infty, \infty)$.

21. (*Calculus required*) Use the Wronskian to show that the functions $f_1(x) = \sin x$, $f_2(x) = \cos x$, and $f_3(x) = x\cos x$ are linearly independent vectors in $C^\infty(-\infty, \infty)$.

22. Show that for any vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in a vector space V, the vectors $\mathbf{u} - \mathbf{v}$, $\mathbf{v} - \mathbf{w}$, and $\mathbf{w} - \mathbf{u}$ form a linearly dependent set.

23. a. In Example 1 we showed that the mutually orthogonal vectors $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$ form a linearly independent set of vectors in R^3. Do you think that every set of three nonzero mutually orthogonal vectors in R^3 is linearly independent? Justify your conclusion with a geometric argument.

b. Justify your conclusion with an algebraic argument. [*Hint:* Use dot products.]

Working with Proofs

24. Prove that if $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly independent set of vectors, then so are $\{\mathbf{v}_1, \mathbf{v}_2\}$, $\{\mathbf{v}_1, \mathbf{v}_3\}$, $\{\mathbf{v}_2, \mathbf{v}_3\}$, $\{\mathbf{v}_1\}$, $\{\mathbf{v}_2\}$, and $\{\mathbf{v}_3\}$.

25. Prove that if $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a linearly independent set of vectors, then so is every nonempty subset of S.

26. Prove that if $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly dependent set of vectors in a vector space V, and $\mathbf{v}_4$ is any vector in V that is not in S, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is also linearly dependent.

27. Prove that if $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a linearly dependent set of vectors in a vector space V, and if $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ are any vectors in V that are not in S, then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ is also linearly dependent.

28. Prove that in P_2 every set with more than three vectors is linearly dependent.

29. Prove that if $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent and $\mathbf{v}_3$ does not lie in span$\{\mathbf{v}_1, \mathbf{v}_2\}$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent.

30. Prove Theorem 4.4.1 in the case where S has only one vector.

31. Prove part (b) of Theorem 4.4.2.

True-False Exercises

TF. In parts **(a)–(h)** determine whether the statement is true or false, and justify your answer.

a. A set containing a single vector is linearly independent.

b. No linearly independent set contains the zero vector.

c. Every linearly dependent set contains the zero vector.

d. If the set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent, then $\{k\mathbf{v}_1, k\mathbf{v}_2, k\mathbf{v}_3\}$ is also linearly independent for every nonzero scalar k.

e. If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly dependent nonzero vectors, then at least one vector $\mathbf{v}_k$ is a unique linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}$.

f. The set of 2×2 matrices that contain exactly two 1's and two 0's is a linearly independent set in M_{22}.

g. The three polynomials $(x-1)(x+2)$, $x(x+2)$, and $x(x-1)$ are linearly independent.

h. The functions f_1 and f_2 are linearly dependent if there is a real number x such that $k_1 f_1(x) + k_2 f_2(x) = 0$ for some scalars k_1 and k_2.

Working with Technology

T1. Devise three different methods for using your technology utility to determine whether a set of vectors in R^n is linearly independent, and then use each of those methods to determine whether the following vectors are linearly independent.

$$\mathbf{v}_1 = (4, -5, 2, 6), \quad \mathbf{v}_2 = (2, -2, 1, 3),$$
$$\mathbf{v}_3 = (6, -3, 3, 9), \quad \mathbf{v}_4 = (4, -1, 5, 6)$$

T2. Show that $S = \{\cos t, \sin t, \cos 2t, \sin 2t\}$ is a linearly independent set in $C(-\infty, \infty)$ by evaluating the left side of the equation

$$c_1 \cos t + c_2 \sin t + c_3 \cos 2t + c_4 \sin 2t = 0$$

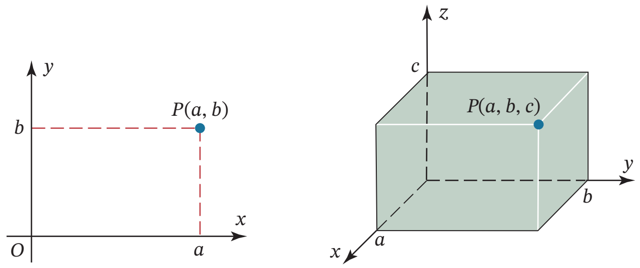
at sufficiently many values of t to obtain a linear system whose only solution is $c_1 = c_2 = c_3 = c_4 = 0$.

4.5 Coordinates and Basis

We usually think of a line as being one-dimensional, a plane as two-dimensional, and the space around us as three-dimensional. It is the primary goal of this section and the next to make this intuitive notion of dimension precise. In this section we will discuss coordinate systems in general vector spaces and lay the groundwork for a precise definition of dimension in the next section.

Coordinate Systems in Linear Algebra

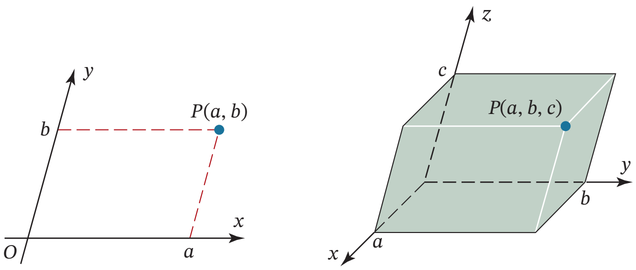
In analytic geometry one uses *rectangular* coordinate systems to create a one-to-one correspondence between points in 2-space and ordered pairs of real numbers and between points in 3-space and ordered triples of real numbers (**Figure 4.5.1**). Although rectangular coordinate systems are common, they are not essential. For example, **Figure 4.5.2** shows coordinate systems in 2-space and 3-space in which the coordinate axes are not mutually perpendicular.



Coordinates of P in a rectangular coordinate system in 2-space.

Coordinates of P in a rectangular coordinate system in 3-space.

FIGURE 4.5.1



Coordinates of P in a nonrectangular coordinate system in 2-space.

Coordinates of P in a nonrectangular coordinate system in 3-space.

FIGURE 4.5.2

In linear algebra coordinate systems are commonly specified using vectors rather than coordinate axes. For example, in **Figure 4.5.3** we have re-created the coordinate systems in Figure 4.5.2 by using unit vectors to identify the positive directions and then attaching coordinates to a point P using the scalar coefficients in the equations

$$\overrightarrow{OP} = a\mathbf{u}_1 + b\mathbf{u}_2 \quad \text{and} \quad \overrightarrow{OP} = a\mathbf{u}_1 + b\mathbf{u}_2 + c\mathbf{u}_3$$

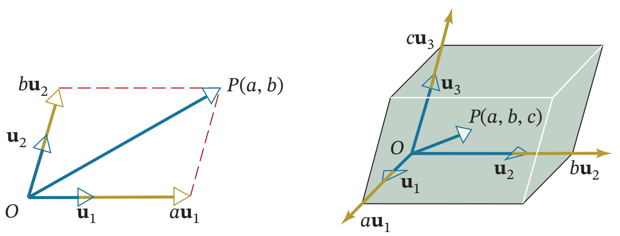


FIGURE 4.5.3

Units of measurement are essential ingredients of any coordinate system. In geometry problems one tries to use the same unit of measurement on all axes to avoid distorting the shapes of figures. This is less important in applications where coordinates represent physical quantities with diverse units (for example, time in seconds on one axis and temperature in degrees Celsius on another axis). To allow for this level of generality, we will relax the requirement that *unit* vectors be used to identify the positive directions and require only that those vectors be linearly independent. We will refer to these as the "basis vectors" for the coordinate system. In summary, it is the directions of the basis vectors that establish the positive directions, and it is the lengths of the basis vectors that establish the spacing between the integer points on the axes (**Figure 4.5.4**).

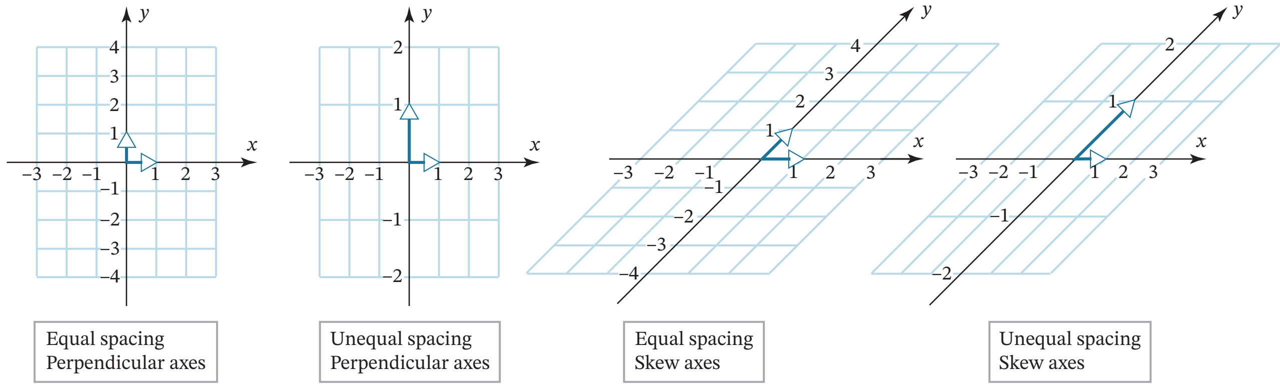


FIGURE 4.5.4

Basis for a Vector Space

Our next goal is to extend the concepts of "basis vectors" and "coordinate systems" to general vector spaces, and for that purpose we will need some definitions. Vector spaces fall into two categories: A vector space V is said to be ***finite-dimensional*** if there is a finite set of vectors in V that spans V and is said to be ***infinite-dimensional*** if no such set exists.

Definition 1

If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a set of vectors in a finite-dimensional vector space V, then S is called a ***basis*** for V if:

(a) S spans V.

(b) S is linearly independent.

If you think of a basis as describing a coordinate system for a finite-dimensional vector space V, then part (a) of this definition guarantees that there are enough basis vectors to provide coordinates for all vectors in V, and part (b) guarantees that there is no interrelationship between the basis vectors. Here are some examples.

EXAMPLE 1 | The Standard Basis for R^n

Recall from Example 1 of Section 4.3 that the standard unit vectors

$$\mathbf{e}_1 = (1, 0, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, 0, \ldots, 0), \ldots, \quad \mathbf{e}_n = (0, 0, 0, \ldots, 1)$$

span R^n and from Example 1 of Section 4.4 that they are linearly independent. Thus, they form a basis for R^n that we call the ***standard basis for R^n***. In particular,

$$\mathbf{i} = (1, 0), \quad \mathbf{j} = (0, 1)$$

and

$$\mathbf{i} = (1, 0, 0), \quad \mathbf{j} = (0, 1, 0), \quad \mathbf{k} = (0, 0, 1)$$

are the standard bases for R^2 and R^3, respectively.

EXAMPLE 2 | The Standard Basis for P_n

Show that $S = \{1, x, x^2, \ldots, x^n\}$ is a basis for the vector space P_n of polynomials of degree n or less.

Solution We must show that the polynomials in S are linearly independent and span P_n. Let us denote these polynomials by

$$\mathbf{p}_0 = 1, \quad \mathbf{p}_1 = x, \quad \mathbf{p}_2 = x^2, \ldots, \quad \mathbf{p}_n = x^n$$

We showed in Example 3 of Section 4.3 that these vectors span P_n and in Example 4 of Section 4.4 that they are linearly independent. Thus, they form a basis for P_n that we call the ***standard basis for P_n***.

EXAMPLE 3 | Another Basis for R^3

Show that the vectors $\mathbf{v}_1 = (1, 2, 1), \mathbf{v}_2 = (2, 9, 0)$, and $\mathbf{v}_3 = (3, 3, 4)$ form a basis for R^3.

Solution We must show that these vectors are linearly independent and span R^3. To prove linear independence we must show that the vector equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{0} \tag{1}$$

has only the trivial solution; and to prove that the vectors span R^3 we must show that every vector $\mathbf{b} = (b_1, b_2, b_3)$ in R^3 can be expressed as

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{b} \tag{2}$$

By equating corresponding components on the two sides, these two equations can be expressed as the linear systems

$$\begin{aligned} c_1 + 2c_2 + 3c_3 &= 0 \\ 2c_1 + 9c_2 + 3c_3 &= 0 \\ c_1 \qquad + 4c_3 &= 0 \end{aligned} \quad \text{and} \quad \begin{aligned} c_1 + 2c_2 + 3c_3 &= b_1 \\ 2c_1 + 9c_2 + 3c_3 &= b_2 \\ c_1 \qquad + 4c_3 &= b_3 \end{aligned} \tag{3}$$

(verify). Thus, we have reduced the problem to showing that in (3) the homogeneous system has only the trivial solution and that the nonhomogeneous system is consistent for all values of b_1, b_2, and b_3. But the two systems have the same coefficient matrix

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

so it follows from parts (*b*), (*e*), and (*g*) of Theorem 2.3.8 that we can prove both results at the same time by showing that $\det(A) \neq 0$. We leave it for you to confirm that $\det(A) = -1$, which proves that the vectors $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$ form a basis for R^3.

From Examples 1 and 3 you can see that a vector space can have more than one basis.

EXAMPLE 4 | The Standard Basis for M_{mn}

Show that the matrices

$$M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad M_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad M_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

form a basis for the vector space M_{22} of 2×2 matrices.

Solution We must show that the matrices are linearly independent and span M_{22}. To prove linear independence we must show that the equation

$$c_1M_1 + c_2M_2 + c_3M_3 + c_4M_4 = \mathbf{0} \tag{4}$$

has only the trivial solution, where $\mathbf{0}$ is the 2×2 zero matrix; and to prove that the matrices span M_{22} we must show that every 2×2 matrix

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

can be expressed as

$$c_1M_1 + c_2M_2 + c_3M_3 + c_4M_4 = B \tag{5}$$

The matrix forms of Equations (4) and (5) are

$$c_1\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + c_2\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c_3\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + c_4\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

and

$$c_1\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + c_2\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c_3\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + c_4\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

which can be rewritten as

$$\begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Since the first equation has only the trivial solution

$$c_1 = c_2 = c_3 = c_4 = 0$$

the matrices are linearly independent, and since the second equation has the solution

$$c_1 = a, \quad c_2 = b, \quad c_3 = c, \quad c_4 = d$$

the matrices span M_{22}. This proves that the matrices M_1, M_2, M_3, M_4 form a basis for M_{22}. More generally, the mn different matrices whose entries are zero except for a single entry of 1 form a basis for M_{mn} called the ***standard basis for*** $\boldsymbol{M_{mn}}$.

The simplest of all vector spaces is the zero vector space $V = \{\mathbf{0}\}$. This space is finite-dimensional because it is spanned by the vector $\mathbf{0}$. However, it has no basis in the sense of Definition 1 because $\{\mathbf{0}\}$ is not a linearly independent set (why?). However, we will find it useful to define the empty set Ø to be a basis for this vector space.

EXAMPLE 5 | An Infinite-Dimensional Vector Space

Show that the vector space of P_∞ of all polynomials with real coefficients is infinite-dimensional by showing that it has no finite spanning set.

Solution If there were a finite spanning set, say $S = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_r\}$, then the degrees of the polynomials in S would have a maximum value, say n; and this in turn would imply that any linear combination of the polynomials in S would have degree at most n. Thus, there would be no way to express the polynomial x^{n+1} as a linear combination of the polynomials in S, contradicting the fact that the vectors in S span P_∞.

EXAMPLE 6 | Some Finite- and Infinite-Dimensional Spaces

In Examples 1, 2, and 4 we found bases for R^n, P_n, and M_{mn}, so these vector spaces are finite-dimensional. We showed in Example 5 that the vector space P_∞ is not spanned by finitely many vectors and hence is infinite-dimensional. Some other examples of infinite-dimensional vector spaces are R^∞, $F(-\infty, \infty)$, $C(-\infty, \infty)$, $C^m(-\infty, \infty)$, and $C^\infty(-\infty, \infty)$.

Coordinates Relative to a Basis

Earlier in this section we drew an informal analogy between basis vectors and coordinate systems. Our next goal is to make this informal idea precise by defining the notion of a coordinate system in a general vector space. The following theorem will be our first step in that direction.

Theorem 4.5.1

Uniqueness of Basis Representation

If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for a vector space V, then every vector $\mathbf{v}$ in V can be expressed in the form $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$ in exactly one way.

Proof Since S spans V, it follows from the definition of a spanning set that every vector in V is expressible as a linear combination of the vectors in S. To see that there is only *one* way to express a vector as a linear combination of the vectors in S, suppose that some vector $\mathbf{v}$ can be written as

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

and also as

$$\mathbf{v} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n$$

Subtracting the second equation from the first gives

$$\mathbf{0} = (c_1 - k_1)\mathbf{v}_1 + (c_2 - k_2)\mathbf{v}_2 + \cdots + (c_n - k_n)\mathbf{v}_n$$

Since the right side of this equation is a linear combination of vectors in S, the linear independence of S implies that

$$c_1 - k_1 = 0, \quad c_2 - k_2 = 0, \ldots, \quad c_n - k_n = 0$$

that is,

$$c_1 = k_1, \quad c_2 = k_2, \ldots, \quad c_n = k_n$$

Thus, the two expressions for $\mathbf{v}$ are the same. ■

We now have all of the ingredients required to define the notion of “coordinates” in a general vector space V. For motivation, observe that in R^3, for example, the coordinates (a, b, c) of a vector $\mathbf{v}$ are precisely the coefficients in the formula

$$\mathbf{v} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$$

that expresses $\mathbf{v}$ as a linear combination of the standard basis vectors for R^3 (see **Figure 4.5.5**).

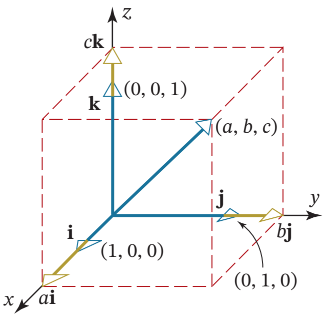


FIGURE 4.5.5

Our next definition will generalize this idea, but first we need to make some observations about bases. Up to now the order of the vectors in a basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ for a vector space V did not matter. The only requirement was that the vectors in the set S be linearly independent and span V. However, in many cases the order in which the vectors in S are listed matters. A basis in which the listed order of the vectors matters is called an ***ordered basis***. Thus, for example, if $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for a vector space V, then $S' = \{\mathbf{v}_2, \mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is also a basis, but it is a different ordered basis.

Definition 2

If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an ordered basis for a vector space V, and

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

is the expression for a vector $\mathbf{v}$ in terms of the basis S, then the scalars $c_1, c_2, \ldots, c_n$ are called the ***coordinates of* v *relative to the basis S***. The vector $(c_1, c_2, \ldots, c_n)$ in R^n constructed from these coordinates is called the ***coordinate vector of* v *relative to S***; it is denoted by

$$(\mathbf{v})_S = (c_1, c_2, \ldots, c_n) \tag{6}$$

Frequently, we will want to express (6) as a column matrix, in which case we will use the notation

$$[\mathbf{v}]_S = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

We call this the ***matrix form*** of the coordinate vector and (6) the ***comma-delimited form***.

Observe that $(\mathbf{v})_S$ is a vector in R^n, so that once an ordered basis S is given for a vector space V, Theorem 4.5.1 establishes a one-to-one correspondence between vectors in V and vectors in R^n (**Figure 4.5.6**).

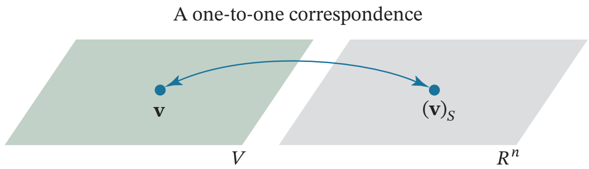


FIGURE 4.5.6

EXAMPLE 7 | Coordinates Relative to the Standard Basis for R^n

In the special case where $V = R^n$ and S is the *standard basis*, the coordinate vector $(\mathbf{v})_S$ and the vector $\mathbf{v}$ are the same; that is,

$$\mathbf{v} = (\mathbf{v})_S$$

For example, in R^3 the representation of a vector $\mathbf{v} = (a, b, c)$ as a linear combination of the vectors in the standard basis $S = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is

$$\mathbf{v} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$$

so the coordinate vector relative to this basis is $(\mathbf{v})_S = (a, b, c)$, which is the same as the vector $\mathbf{v}$.

EXAMPLE 8 | Coordinate Vectors Relative to Standard Bases

(*a*) Find the coordinate vector for the polynomial

$$\mathbf{p}(x) = c_0 + c_1x + c_2x^2 + \cdots + c_nx^n$$

relative to the standard basis for the vector space P_n.

(*b*) Find the coordinate vector of

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

relative to the standard basis for M_{22}.

Solution (*a*) The given formula for $\mathbf{p}(x)$ expresses this polynomial as a linear combination of the standard basis vectors $S = \{1, x, x^2, \ldots, x^n\}$. Thus, the coordinate vector for $\mathbf{p}$ relative to S is

$$(\mathbf{p})_S = (c_0, c_1, c_2, \ldots, c_n)$$

Solution (*b*) We showed in Example 4 that the representation of a vector

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

as a linear combination of the standard basis vectors is

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

so the coordinate vector of B relative to S is

$$(B)_S = (a, b, c, d)$$

EXAMPLE 9 | Coordinates in R^3

(*a*) We showed in Example 3 that the vectors

$$\mathbf{v}_1 = (1, 2, 1), \quad \mathbf{v}_2 = (2, 9, 0), \quad \mathbf{v}_3 = (3, 3, 4)$$

form a basis for R^3. Find the coordinate vector of $\mathbf{v} = (5, -1, 9)$ relative to the basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

(*b*) Find the vector $\mathbf{v}$ in R^3 whose coordinate vector relative to S is $(\mathbf{v})_S = (-1, 3, 2)$.

Solution (*a*) To find $(\mathbf{v})_S$ we must first express $\mathbf{v}$ as a linear combination of the vectors in S; that is, we must find values of c_1, c_2, and c_3 such that

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$$

or, in terms of components,

$$(5, -1, 9) = c_1(1, 2, 1) + c_2(2, 9, 0) + c_3(3, 3, 4)$$

Equating corresponding components gives

$$\begin{aligned} c_1 + 2c_2 + 3c_3 &= 5 \\ 2c_1 + 9c_2 + 3c_3 &= -1 \\ c_1 \qquad\; + 4c_3 &= 9 \end{aligned}$$

Solving this system we obtain $c_1 = 1$, $c_2 = -1$, $c_3 = 2$ (verify). Therefore,

$$(\mathbf{v})_S = (1, -1, 2)$$

Solution (*b*) Using the definition of $(\mathbf{v})_S$, we obtain

$$\begin{aligned} \mathbf{v} &= (-1)\mathbf{v}_1 + 3\mathbf{v}_2 + 2\mathbf{v}_3 \\ &= (-1)(1, 2, 1) + 3(2, 9, 0) + 2(3, 3, 4) = (11, 31, 7) \end{aligned}$$

Exercise Set 4.5

1. Use the method of Example 3 to show that the following set of vectors forms a basis for R^2.

$$\{(2,1),(3,0)\}$$

2. Use the method of Example 3 to show that the following set of vectors forms a basis for R^3.

$$\{(3,1,-4),(2,5,6),(1,4,8)\}$$

3. Show that the following polynomials form a basis for P_2.

$$x^2+1, \quad x^2-1, \quad 2x-1$$

4. Show that the following polynomials form a basis for P_3.

$$1+x, \quad 1-x, \quad 1-x^2, \quad 1-x^3$$

5. Show that the following matrices form a basis for M_{22}.

$$\begin{bmatrix} 3 & 6 \\ 3 & -6 \end{bmatrix}, \quad \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & -8 \\ -12 & -4 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$$

6. Show that the following matrices form a basis for M_{22}.

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

7. In each part, show that the set of vectors is not a basis for R^3.

   a. $\{(2,-3,1),(4,1,1),(0,-7,1)\}$

   b. $\{(1,6,4),(2,4,-1),(-1,2,5)\}$

8. Show that the following vectors do not form a basis for P_2.

$$1-3x+2x^2, \quad 1+x+4x^2, \quad 1-7x$$

9. Show that the following matrices do not form a basis for M_{22}.

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

10. Let V be the space spanned by $\mathbf{v}_1=\cos^2 x$, $\mathbf{v}_2=\sin^2 x$, $\mathbf{v}_3=\cos 2x$.

    a. Show that $S=\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\}$ is not a basis for V.

    b. Find a basis for V.

11. Find the coordinate vector of $\mathbf{w}$ relative to the basis $S=\{\mathbf{u}_1,\mathbf{u}_2\}$ for R^2.

    a. $\mathbf{u}_1=(2,-4)$, $\mathbf{u}_2=(3,8)$; $\mathbf{w}=(1,1)$

    b. $\mathbf{u}_1=(1,1)$, $\mathbf{u}_2=(0,2)$; $\mathbf{w}=(a,b)$

12. Find the coordinate vector of $\mathbf{w}$ relative to the basis $S=\{\mathbf{u}_1,\mathbf{u}_2\}$ for R^2.

    a. $\mathbf{u}_1=(1,-1)$, $\mathbf{u}_2=(1,1)$; $\mathbf{w}=(1,0)$

    b. $\mathbf{u}_1=(1,-1)$, $\mathbf{u}_2=(1,1)$; $\mathbf{w}=(0,1)$

13. Find the coordinate vector of $\mathbf{v}$ relative to the basis $S=\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\}$ for R^3.

    a. $\mathbf{v}=(2,-1,3)$; $\mathbf{v}_1=(1,0,0)$, $\mathbf{v}_2=(2,2,0)$, $\mathbf{v}_3=(3,3,3)$

    b. $\mathbf{v}=(5,-12,3)$; $\mathbf{v}_1=(1,2,3)$, $\mathbf{v}_2=(-4,5,6)$, $\mathbf{v}_3=(7,-8,9)$

14. Find the coordinate vector of $\mathbf{p}$ relative to the basis $S=\{\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3\}$ for P_2.

    a. $\mathbf{p}=4-3x+x^2$; $\mathbf{p}_1=1$, $\mathbf{p}_2=x$, $\mathbf{p}_3=x^2$

    b. $\mathbf{p}=2-x+x^2$; $\mathbf{p}_1=1+x$, $\mathbf{p}_2=1+x^2$, $\mathbf{p}_3=x+x^2$

In Exercises 15–16, first show that the set $S=\{A_1,A_2,A_3,A_4\}$ is a basis for M_{22}, then express A as a linear combination of the vectors in S, and then find the coordinate vector of A relative to S.

15. $A_1=\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $A_2=\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$, $A_3=\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$, $A_4=\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$; $A=\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$

16. $A_1=\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, $A_2=\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, $A_3=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $A_4=\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$; $A=\begin{bmatrix} 6 & 2 \\ 5 & 3 \end{bmatrix}$

In Exercises 17–18, first show that the set $S=\{\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3\}$ is a basis for P_2, then express $\mathbf{p}$ as a linear combination of the vectors in S, and then find the coordinate vector of $\mathbf{p}$ relative to S.

17. $\mathbf{p}_1=1+x+x^2$, $\mathbf{p}_2=x+x^2$, $\mathbf{p}_3=x^2$; $\mathbf{p}=7-x+2x^2$

18. $\mathbf{p}_1=1+2x+x^2$, $\mathbf{p}_2=2+9x$, $\mathbf{p}_3=3+3x+4x^2$; $\mathbf{p}=2+17x-3x^2$

19. In words, explain why the sets of vectors in parts (a) to (d) are *not* bases for the indicated vector spaces.

    a. $\mathbf{u}_1=(1,2)$, $\mathbf{u}_2=(0,3)$, $\mathbf{u}_3=(1,5)$ for R^2

    b. $\mathbf{u}_1=(-1,3,2)$, $\mathbf{u}_2=(6,1,1)$ for R^3

    c. $\mathbf{p}_1=1+x+x^2$, $\mathbf{p}_2=x$ for P_2

    d. $A=\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}$, $B=\begin{bmatrix} 6 & 0 \\ -1 & 4 \end{bmatrix}$, $C=\begin{bmatrix} 3 & 0 \\ 1 & 7 \end{bmatrix}$, $D=\begin{bmatrix} 5 & 0 \\ 4 & 2 \end{bmatrix}$ for M_{22}

20. In any vector space a set that contains the zero vector must be linearly dependent. Explain why this is so.

21. In each part, let $T_A: R^3 \to R^3$ be multiplication by A, and let $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ be the standard basis for R^3. Determine whether the set $\{T_A(\mathbf{e}_1),T_A(\mathbf{e}_2),T_A(\mathbf{e}_3)\}$ is linearly independent in R^2.

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

22. In each part, let $T_A: R^3 \to R^3$ be multiplication by A, and let $\mathbf{u}=(1,-2,-1)$. Find the coordinate vector of $T_A(\mathbf{u})$ relative to the basis $S=\{(1,1,0),(0,1,1),(1,1,1)\}$ for R^3.

    a. $A=\begin{bmatrix} 2 & -1 & 0 \\ 1 & 1 & 1 \\ 0 & -1 & 2 \end{bmatrix}$ **b.** $A=\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

23. The accompanying figure shows a rectangular xy-coordinate system determined by the unit basis vectors $\mathbf{i}$ and $\mathbf{j}$ and an $x'y'$-coordinate system determined by unit basis vectors $\mathbf{u}_1$ and $\mathbf{u}_2$. Find the $x'y'$-coordinates of the points whose xy-coordinates are given.

a. $(\sqrt{3}, 1)$ **b.** $(1, 0)$ **c.** $(0, 1)$ **d.** (a, b)

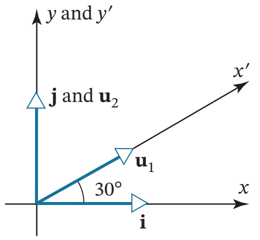


FIGURE Ex-23

24. The accompanying figure shows a rectangular xy-coordinate system and an $x'y'$-coordinate system with skewed axes. Assuming that 1-unit scales are used on all the axes, find the $x'y'$-coordinates of the points whose xy-coordinates are given.

a. $(1, 1)$ **b.** $(1, 0)$ **c.** $(0, 1)$ **d.** (a, b)

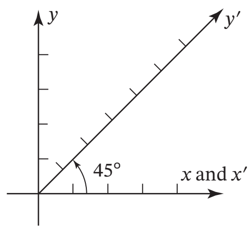


FIGURE Ex-24

25. The first four ***Hermite polynomials*** [named for the French mathematician Charles Hermite (1822–1901)] are

$$1, \quad 2t, \quad -2 + 4t^2, \quad -12t + 8t^3$$

These polynomials have a wide variety of applications in physics and engineering.

a. Show that the first four Hermite polynomials form a basis for P_3.

b. Let B be the basis in part (a). Find the coordinate vector of the polynomial

$$\mathbf{p}(t) = -1 - 4t + 8t^2 + 8t^3$$

relative to B.

26. The first four ***Laguerre polynomials*** [named for the French mathematician Edmond Laguerre (1834–1886)] are

$$1, \quad 1 - t, \quad 2 - 4t + t^2, \quad 6 - 18t + 9t^2 - t^3$$

a. Show that the first four Laguerre polynomials form a basis for P_3.

b. Let B be the basis in part (a). Find the coordinate vector of the polynomial

$$\mathbf{p}(t) = -10t + 9t^2 - t^3$$

relative to B.

27. Consider the coordinate vectors

$$[\mathbf{w}]_S = \begin{bmatrix} 6 \\ -1 \\ 4 \end{bmatrix}, \quad [\mathbf{q}]_S = \begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix}, \quad [B]_S = \begin{bmatrix} -8 \\ 7 \\ 6 \\ 3 \end{bmatrix}$$

a. Find $\mathbf{w}$ if S is the basis in Exercise 2.

b. Find $\mathbf{q}$ if S is the basis in Exercise 3.

c. Find B if S is the basis in Exercise 5.

28. The basis that we gave for M_{22} in Example 4 consisted of noninvertible matrices. Do you think that there is a basis for M_{22} consisting of invertible matrices? Justify your answer.

Working with Proofs

29. Prove that R^∞ is an infinite-dimensional vector space.

30. Let $T_A: R^n \to R^n$ be multiplication by an invertible matrix A, and let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a basis for R^n. Prove that $\{T_A(\mathbf{u}_1), T_A(\mathbf{u}_2), \ldots, T_A(\mathbf{u}_n)\}$ is also a basis for R^n.

31. Prove that if V is a subspace of a vector space W and if V is infinite-dimensional, then so is W.

True-False Exercises

TF. In parts (a)–(e) determine whether the statement is true or false, and justify your answer.

a. If $V = \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, then $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for V.

b. Every linearly independent subset of a vector space V is a basis for V.

c. If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for a vector space V, then every vector in V can be expressed as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.

d. The coordinate vector of a vector $\mathbf{x}$ in R^n relative to the standard basis for R^n is $\mathbf{x}$.

e. Every basis of P_4 contains at least one polynomial of degree 3 or less.

Working with Technology

T1. Let V be the subspace of P_3 spanned by the vectors

$$\mathbf{p}_1 = 1 + 5x - 3x^2 - 11x^3, \quad \mathbf{p}_2 = 7 + 4x - x^2 + 2x^3,$$
$$\mathbf{p}_3 = 5 + x + 9x^2 + 2x^3, \quad \mathbf{p}_4 = 3 - x + 7x^2 + 5x^3$$

a. Find a basis S for V.

b. Find the coordinate vector of $\mathbf{p} = 19 + 18x - 13x^2 - 10x^3$ relative to the basis S you obtained in part (a).

T2. Let V be the subspace of $C^\infty(-\infty, \infty)$ spanned by the vectors in the set

$$B = \{1, \cos x, \cos^2 x, \cos^3 x, \cos^4 x, \cos^5 x\}$$

and accept without proof that B is a basis for V. Confirm that the following vectors are in V, and find their coordinate vectors relative to B.

$$\mathbf{f}_0 = 1, \quad \mathbf{f}_1 = \cos x, \quad \mathbf{f}_2 = \cos 2x, \quad \mathbf{f}_3 = \cos 3x,$$
$$\mathbf{f}_4 = \cos 4x, \quad \mathbf{f}_5 = \cos 5x$$

4.6 Dimension

We showed in the previous section that the standard basis for R^n has n vectors and hence that the standard basis for R^3 has three vectors, the standard basis for R^2 has two vectors, and the standard basis for R^1 $(= R)$ has one vector. Since we think of space as three-dimensional, a plane as two-dimensional, and a line as one-dimensional, there seems to be a link between the number of vectors in a basis and the dimension of a vector space. We will develop this idea in this section.

Number of Vectors in a Basis

Our first goal in this section is to establish the following fundamental theorem.

Theorem 4.6.1

All bases for a finite-dimensional vector space have the same number of vectors.

To prove this theorem we will need the following preliminary result, whose proof is deferred to the end of the section.

Theorem 4.6.2

Let V be a finite-dimensional vector space, and let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be any basis for V.

(a) If a set in V has more than n vectors, then it is linearly dependent.

(b) If a set in V has fewer than n vectors, then it does not span V.

We can now see rather easily why Theorem 4.6.1 is true; for if

$$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

is an *arbitrary* basis for V, then the linear independence of S implies that any set in V with more than n vectors is linearly dependent and any set in V with fewer than n vectors does not span V. Thus, unless a set in V has exactly n vectors it cannot be a basis.

We noted in the introduction to this section that for certain familiar vector spaces the intuitive notion of dimension coincides with the number of vectors in a basis. The following definition makes this idea precise.

Engineers often use the term ***degrees of freedom*** as a synonym for dimension.

Definition 1

The ***dimension*** of a finite-dimensional vector space V is denoted by $\dim(V)$ and is defined to be the number of vectors in a basis for V. In addition, the zero vector space is defined to have dimension zero.

EXAMPLE 1 | Dimensions of Some Familiar Vector Spaces

$$\begin{aligned}
\dim(R^n) &= n && \textbf{[The standard basis has } \boldsymbol{n} \textbf{ vectors.]}\\
\dim(P_n) &= n+1 && \textbf{[The standard basis has } \boldsymbol{n+1} \textbf{ vectors.]}\\
\dim(M_{mn}) &= mn && \textbf{[The standard basis has } \boldsymbol{mn} \textbf{ vectors.]}
\end{aligned}$$

EXAMPLE 2 | Dimension of Span(S)

If $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ then every vector in span(S) is expressible as a linear combination of the vectors in S. Thus, if the vectors in S are *linearly independent*, they automatically form a basis for span(S), from which we can conclude that

$$\dim\left[\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}\right] = r$$

In words, the dimension of the space spanned by a linearly independent set of vectors is equal to the number of vectors in that set.

EXAMPLE 3 | Dimension of a Solution Space

Find a basis for and the dimension of the solution space of the homogeneous system

$$\begin{aligned}
x_1 + 3x_2 - 2x_3 \phantom{{}- 2x_4} + 2x_5 \phantom{{}- 3x_6} &= 0 \\
2x_1 + 6x_2 - 5x_3 - 2x_4 + 4x_5 - 3x_6 &= 0 \\
5x_3 + 10x_4 \phantom{{}+ 4x_5} + 15x_6 &= 0 \\
2x_1 + 6x_2 \phantom{{}- 5x_3} + 8x_4 + 4x_5 + 18x_6 &= 0
\end{aligned}$$

Solution In Example 6 of Section 1.2 we found the solution of this system to be

$$x_1 = -3r - 4s - 2t, \quad x_2 = r, \quad x_3 = -2s, \quad x_4 = s, \quad x_5 = t, \quad x_6 = 0$$

which can be written in vector form as

$$(x_1, x_2, x_3, x_4, x_5, x_6) = (-3r - 4s - 2t, r, -2s, s, t, 0)$$

or, alternatively, as

$$(x_1, x_2, x_3, x_4, x_5, x_6) = r(-3, 1, 0, 0, 0, 0) + s(-4, 0, -2, 1, 0, 0) + t(-2, 0, 0, 0, 1, 0)$$

This shows that the vectors

$$\mathbf{v}_1 = (-3, 1, 0, 0, 0, 0), \quad \mathbf{v}_2 = (-4, 0, -2, 1, 0, 0), \quad \mathbf{v}_3 = (-2, 0, 0, 0, 1, 0)$$

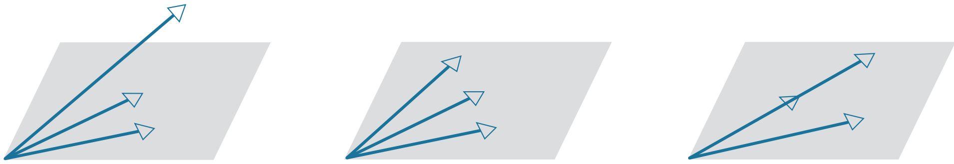
span the solution space. We leave it for you to check that these vectors are linearly independent by showing that none of them is a linear combination of the other two (but see the remark that follows). Thus, the solution space has dimension 3.

Remark It can be shown that for any homogeneous linear system, the method of the last example *always* produces a basis for the solution space of the system. We omit the formal proof.

Some Fundamental Theorems

We will devote the remainder of this section to a series of theorems that reveal the subtle interrelationships among the concepts of linear independence, spanning sets, basis, and dimension. These theorems are not simply exercises in mathematical theory—they are essential to the understanding of vector spaces and the applications that build on them.

We will start with a theorem (proved at the end of this section) that is concerned with the effect on linear independence and spanning if a vector is added to or removed from a nonempty set of vectors. Informally stated, if you start with a linearly independent set S and adjoin to it a vector that is not a linear combination of those already in S, then the enlarged set will still be linearly independent. Also, if you start with a set S of two or more vectors in which one of the vectors is a linear combination of the others, then that vector can be removed from S without affecting span(S) (**Figure 4.6.1**).



The vector outside the plane can be adjoined to the other two without affecting their linear independence.

Any of the vectors can be removed, and the remaining two will still span the plane.

Either of the collinear vectors can be removed, and the remaining two will still span the plane.

FIGURE 4.6.1

Theorem 4.6.3

Plus/Minus Theorem

Let S be a nonempty set of vectors in a vector space V.

(*a*) If S is a linearly independent set, and if $\mathbf{v}$ is a vector in V that is outside of span(S), then the set $S \cup \{\mathbf{v}\}$ that results by inserting $\mathbf{v}$ into S is still linearly independent.

(*b*) If $\mathbf{v}$ is a vector in S that is expressible as a linear combination of other vectors in S, and if $S - \{\mathbf{v}\}$ denotes the set obtained by removing $\mathbf{v}$ from S, then S and $S - \{\mathbf{v}\}$ span the same space; that is,

$$\text{span}(S) = \text{span}\big(S - \{\mathbf{v}\}\big)$$

EXAMPLE 4 | Applying the Plus/Minus Theorem

Show that $\mathbf{p}_1 = 1 - x^2$, $\mathbf{p}_2 = 2 - x^2$, and $\mathbf{p}_3 = x^3$ are linearly independent vectors.

Solution The set $S = \{\mathbf{p}_1, \mathbf{p}_2\}$ is linearly independent since neither vector in S is a scalar multiple of the other. Since the vector $\mathbf{p}_3$ cannot be expressed as a linear combination of the vectors in S (why?), it can be adjoined to S to produce a linearly independent set

$$S \cup \{\mathbf{p}_3\} = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$$

In general, to show that a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for a vector space V, one must show that the vectors are linearly independent and span V. However, if we happen to know that V has dimension n (so that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ contains the right number of vectors for a basis), then it suffices to check *either* linear independence *or* spanning—the remaining condition will hold automatically. This is the content of the following theorem.

Theorem 4.6.4

Let V be an n-dimensional vector space, and let S be a set in V with exactly n vectors. Then S is a basis for V if and only if S spans V or S is linearly independent.

Proof Assume that S has exactly n vectors and spans V. To prove that S is a basis, we must show that S is a linearly independent set. But if this is not so, then some vector $\mathbf{v}$ in S is a linear combination of the remaining vectors. If we remove this vector from S, then it follows from Theorem 4.6.3(b) that the remaining set of $n-1$ vectors still spans V. But this is impossible since Theorem 4.6.2(b) states that no set with fewer than n vectors can span an n-dimensional vector space. Thus S is linearly independent.

Assume that S has exactly n vectors and is a linearly independent set. To prove that S is a basis, we must show that S spans V. But if this is not so, then there is some vector $\mathbf{v}$ in V that is not in span(S). If we insert this vector into S, then it follows from Theorem 4.6.3(a) that this set of $n+1$ vectors is still linearly independent. But this is impossible, since Theorem 4.6.2(a) states that no set with more than n vectors in an n-dimensional vector space can be linearly independent. Thus S spans V. ■

EXAMPLE 5 | Bases by Inspection

(a) Explain why the vectors $\mathbf{v}_1 = (-3, 7)$ and $\mathbf{v}_2 = (5, 5)$ form a basis for R^2.

(b) Explain why the vectors $\mathbf{v}_1 = (2, 0, -1)$, $\mathbf{v}_2 = (4, 0, 7)$, and $\mathbf{v}_3 = (-1, 1, 4)$ form a basis for R^3.

Solution (a) Since neither vector is a scalar multiple of the other, the two vectors form a linearly independent set in the two-dimensional space R^2, and hence they form a basis by Theorem 4.6.4.

Solution (b) The vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ form a linearly independent set in the xz-plane (why?). The vector $\mathbf{v}_3$ is outside of the xz-plane, so the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is also linearly independent. Since R^3 is three-dimensional, Theorem 4.6.4 implies that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis for the vector space R^3.

The next theorem (whose proof is deferred to the end of this section) reveals two important facts about the vectors in a finite-dimensional vector space V:

1. Every spanning set for a subspace is either a basis for that subspace or has a basis as a subset.
2. Every linearly independent set in a subspace is either a basis for that subspace or can be extended to a basis for it.

Theorem 4.6.5

Let S be a finite set of vectors in a finite-dimensional vector space V.

(a) If S spans V but is not a basis for V, then S can be reduced to a basis for V by removing appropriate vectors from S.

(b) If S is a linearly independent set that is not already a basis for V, then S can be enlarged to a basis for V by inserting appropriate vectors into S.

We conclude this section with a theorem that relates the dimension of a vector space to the dimensions of its subspaces.

Theorem 4.6.6

If W is a subspace of a finite-dimensional vector space V, then:

(*a*) W is finite-dimensional.

(*b*) $\dim(W) \leq \dim(V)$.

(*c*) $W = V$ if and only if $\dim(W) = \dim(V)$.

Proof (a) We will leave the proof of this part as an exercise.

Proof (b) Part (*a*) tells us that W is finite-dimensional, so it has a basis

$$S = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$$

Either S is also a basis for V or it is not. If it is a basis, then $\dim(V) = m$, which means that $\dim(V) = \dim(W)$. If not, then because S is a linearly independent set it can be enlarged to a basis for V by part (*b*) of Theorem 4.6.5. But this implies that $\dim(W) < \dim(V)$, so we have shown that $\dim(W) \leq \dim(V)$ in all cases.

Proof (c) Assume that $\dim(W) = \dim(V)$ and that

$$S = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$$

is a basis for W. If S is not also a basis for V, then because it is linearly independent, it can be extended to a basis for V by part (*b*) of Theorem 4.6.5. But this would mean that $\dim(V) > \dim(W)$, which contradicts our hypothesis. Thus S must also be a basis for V, which means that $W = V$. The converse is obvious. ■

Figure 4.6.2 illustrates the geometric relationship between the subspaces of R^3 in order of increasing dimension.

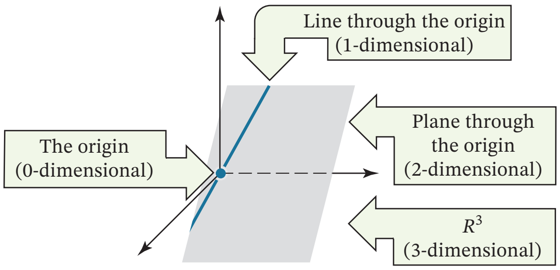


FIGURE 4.6.2

OPTIONAL: We conclude this section with optional proofs of Theorems 4.6.2, 4.6.3, and 4.6.5.

Proof of Theorem 4.6.2(a) Let $S' = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ be any set of m vectors in V, where $m > n$. We want to show that S' is linearly dependent. Since $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis, each $\mathbf{w}_i$ can be expressed as a linear combination of the vectors in S, say

$$\begin{aligned}
\mathbf{w}_1 &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + \cdots + a_{n1}\mathbf{v}_n \\
\mathbf{w}_2 &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + \cdots + a_{n2}\mathbf{v}_n \\
&\ \ \vdots \\
\mathbf{w}_m &= a_{1m}\mathbf{v}_1 + a_{2m}\mathbf{v}_2 + \cdots + a_{nm}\mathbf{v}_n
\end{aligned} \tag{1}$$

To show that S' is linearly dependent, we must find scalars $k_1, k_2, \ldots, k_m$, not all zero, such that

$$k_1\mathbf{w}_1 + k_2\mathbf{w}_2 + \cdots + k_m\mathbf{w}_m = \mathbf{0} \tag{2}$$

We leave it for you to verify that the equations in (1) can be rewritten in the partitioned form

$$[\mathbf{w}_1 \mid \mathbf{w}_2 \mid \cdots \mid \mathbf{w}_m] = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_n] \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} \tag{3}$$

Since $m > n$, the linear system

$$\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{4}$$

has more equations than unknowns and hence has a nontrivial solution

$$x_1 = k_1, \quad x_2 = k_2, \ldots, \quad x_m = k_m$$

Creating a column vector from this solution and multiplying both sides of (3) on the right by this vector yields

$$[\mathbf{w}_1 \mid \mathbf{w}_2 \mid \cdots \mid \mathbf{w}_m] \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_m \end{bmatrix} = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_n] \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_m \end{bmatrix}$$

By (4), this simplifies to

$$[\mathbf{w}_1 \mid \mathbf{w}_2 \mid \cdots \mid \mathbf{w}_m] \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

which we can rewrite as

$$k_1\mathbf{w}_1 + k_2\mathbf{w}_2 + \cdots + k_m\mathbf{w}_m = \mathbf{0}$$

Since the scalar coefficients in this equation are not all zero, we have proved that $S' = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ is linearly independent. ■

The proof of Theorem 4.6.2(b) closely parallels that of Theorem 4.6.2(a) and will be omitted.

Proof of Theorem 4.6.3(a) Assume that $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a linearly independent set in V, and $\mathbf{v}$ is a vector in V that is outside of span(S). To show that $S' = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r, \mathbf{v}\}$ is a linearly independent set, we must show that the only scalars that satisfy

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_r\mathbf{v}_r + k_{r+1}\mathbf{v} = \mathbf{0} \tag{5}$$

are $k_1 = k_2 = \cdots = k_r = k_{r+1} = 0$. But it must be true that $k_{r+1} = 0$ for otherwise we could solve (5) for $\mathbf{v}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$, contradicting the assumption that $\mathbf{v}$ is outside of span(S). Thus, (5) simplifies to

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_r\mathbf{v}_r = \mathbf{0} \tag{6}$$

which, by the linear independence of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$, implies that

$$k_1 = k_2 = \cdots = k_r = 0$$

Proof of Theorem 4.6.3(b) Assume that $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a set of vectors in V, and (to be specific) suppose that $\mathbf{v}_r$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{r-1}$, say

$$\mathbf{v}_r = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_{r-1}\mathbf{v}_{r-1} \tag{7}$$

We want to show that if $\mathbf{v}_r$ is removed from S, then the remaining set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{r-1}\}$ still spans S; that is, we must show that every vector $\mathbf{w}$ in span(S) is expressible as a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{r-1}\}$. But if $\mathbf{w}$ is in span(S), then $\mathbf{w}$ is expressible in the form

$$\mathbf{w} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_{r-1}\mathbf{v}_{r-1} + k_r\mathbf{v}_r$$

or, on substituting (7),

$$\mathbf{w} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_{r-1}\mathbf{v}_{r-1} + k_r(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_{r-1}\mathbf{v}_{r-1})$$

which expresses $\mathbf{w}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{r-1}$.

Proof of Theorem 4.6.5(a) If S is a set of vectors that spans V but is not a basis for V, then S is a linearly dependent set. Thus some vector $\mathbf{v}$ in S is expressible as a linear combination of the other vectors in S. By the Plus/Minus Theorem (4.6.3b), we can remove $\mathbf{v}$ from S, and the resulting set S' will still span V. If S' is linearly independent, then S' is a basis for V, and we are done. If S' is linearly dependent, then we can remove some appropriate vector from S' to produce a set S'' that still spans V. We can continue removing vectors in this way until we finally arrive at a set of vectors in S that is linearly independent and spans V. This subset of S is a basis for V.

Proof of Theorem 4.6.5(b) Suppose that $\dim(V) = n$. If S is a linearly independent set that is not already a basis for V, then S fails to span V, so there is some vector $\mathbf{v}$ in V that is not in span(S). By the Plus/Minus Theorem (4.6.3a), we can insert $\mathbf{v}$ into S, and the resulting set S' will still be linearly independent. If S' spans V, then S' is a basis for V, and we are finished. If S' does not span V, then we can insert an appropriate vector into S' to produce a set S'' that is still linearly independent. We can continue inserting vectors in this way until we reach a set with n linearly independent vectors in V. This set will be a basis for V by Theorem 4.6.4. ■

Exercise Set 4.6

In Exercises 1–6, find a basis for the solution space of the homogeneous linear system, and find the dimension of that space.

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

2. $\begin{aligned} 3x_1 + x_2 + x_3 + x_4 &= 0 \\ 5x_1 - x_2 + x_3 - x_4 &= 0 \end{aligned}$

3. $\begin{aligned} 2x_1 + x_2 + 3x_3 &= 0 \\ x_1 \qquad + 5x_3 &= 0 \\ x_2 + x_3 &= 0 \end{aligned}$

4. $\begin{aligned} x_1 - 4x_2 + 3x_3 - x_4 &= 0 \\ 2x_1 - 8x_2 + 6x_3 - 2x_4 &= 0 \end{aligned}$

5. $\begin{aligned} x_1 - 3x_2 + x_3 &= 0 \\ 2x_1 - 6x_2 + 2x_3 &= 0 \\ 3x_1 - 9x_2 + 3x_3 &= 0 \end{aligned}$

6. $\begin{aligned} x + y + z &= 0 \\ 3x + 2y - 2z &= 0 \\ 4x + 3y - z &= 0 \\ 6x + 5y + z &= 0 \end{aligned}$

7. In each part, find a basis for the given subspace of R^3, and state its dimension.

a. The plane $3x - 2y + 5z = 0$.

b. The plane $x - y = 0$.

c. The line $x = 2t, y = -t, z = 4t$.

d. All vectors of the form (a, b, c), where $b = a + c$.

8. In each part, find a basis for the given subspace of R^4, and state its dimension.

a. All vectors of the form $(a, b, c, 0)$.

b. All vectors of the form (a, b, c, d), where $d = a + b$ and $c = a - b$.

c. All vectors of the form (a, b, c, d), where $a = b = c = d$.

9. Find the dimension of each of the following vector spaces.

a. The vector space of all diagonal $n \times n$ matrices.

b. The vector space of all symmetric $n \times n$ matrices.

c. The vector space of all upper triangular $n \times n$ matrices.

10. Find the dimension of the subspace of P_3 consisting of all polynomials $a_0 + a_1x + a_2x^2 + a_3x^3$ for which $a_0 = 0$.

11. a. Show that the set W of all polynomials in P_2 such that $p(1) = 0$ is a subspace of P_2.

b. Make a conjecture about the dimension of W.

c. Confirm your conjecture by finding a basis for W.

12. Find a standard basis vector for R^3 that can be added to the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ to produce a basis for R^3.

a. $\mathbf{v}_1 = (-1, 2, 3)$, $\mathbf{v}_2 = (1, -2, -2)$

b. $\mathbf{v}_1 = (1, -1, 0)$, $\mathbf{v}_2 = (3, 1, -2)$

13. Find standard basis vectors for R^4 that can be added to the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ to produce a basis for R^4.

$$\mathbf{v}_1 = (1, -4, 2, -3), \quad \mathbf{v}_2 = (-3, 8, -4, 6)$$

14. Let $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ be a basis for a vector space V. Show that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is also a basis, where $\mathbf{u}_1 = \mathbf{v}_1$, $\mathbf{u}_2 = \mathbf{v}_1 + \mathbf{v}_2$, and $\mathbf{u}_3 = \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3$.

15. The vectors $\mathbf{v}_1 = (1, -2, 3)$ and $\mathbf{v}_2 = (0, 5, -3)$ are linearly independent. Enlarge $\{\mathbf{v}_1, \mathbf{v}_2\}$ to a basis for R^3.

16. The vectors $\mathbf{v}_1 = (1, 0, 0, 0)$ and $\mathbf{v}_2 = (1, 1, 0, 0)$ are linearly independent. Enlarge $\{\mathbf{v}_1, \mathbf{v}_2\}$ to a basis for R^4.

17. Find a basis for the subspace of R^3 that is spanned by the vectors

$$\mathbf{v}_1 = (1, 0, 0), \ \mathbf{v}_2 = (1, 0, 1), \ \mathbf{v}_3 = (2, 0, 1), \ \mathbf{v}_4 = (0, 0, -1)$$

18. Find a basis for the subspace of R^4 that is spanned by the vectors

$$\mathbf{v}_1 = (1, 1, 1, 1), \quad \mathbf{v}_2 = (2, 2, 2, 0), \quad \mathbf{v}_3 = (0, 0, 0, 3),$$
$$\mathbf{v}_4 = (3, 3, 3, 4)$$

19. In each part, let $T_A: R^3 \to R^3$ be multiplication by A and find the dimension of the subspace of R^3 consisting of all vectors $\mathbf{x}$ for which $T_A(\mathbf{x}) = \mathbf{0}$.

a. $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 0 \end{bmatrix}$

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

20. In each part, let T_A be multiplication by A and find the dimension of the subspace R^4 consisting of all vectors $\mathbf{x}$ for which $T_A(\mathbf{x}) = \mathbf{0}$.

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

Working with Proofs

21. a. Prove that for every positive integer n, one can find $n + 1$ linearly independent vectors in $F(-\infty, \infty)$. [*Hint:* Look for polynomials.]

b. Use the result in part (a) to prove that $F(-\infty, \infty)$ is infinite-dimensional.

c. Prove that $C(-\infty, \infty)$, $C^m(-\infty, \infty)$, and $C^\infty(-\infty, \infty)$ are infinite-dimensional.

22. Let S be a basis for an n-dimensional vector space V. Prove that if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ form a linearly independent set of vectors in V, then the coordinate vectors $(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, \ldots, (\mathbf{v}_r)_S$ form a linearly independent set in R^n, and conversely.

23. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a nonempty set of vectors in an n-dimensional vector space V. Prove that if the vectors in S span V, then the coordinate vectors $(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, \ldots, (\mathbf{v}_r)_S$ span R^n, and conversely.

24. Prove part (a) of Theorem 4.6.6.

25. Prove: A subspace of a finite-dimensional vector space is finite-dimensional.

26. State the two parts of Theorem 4.6.2 in contrapositive form.

27. In each part, let S be the standard basis for P_2. Use the results proved in Exercises 22 and 23 to find a basis for the subspace of P_2 spanned by the given vectors.

a. $-1 + x - 2x^2, \ 3 + 3x + 6x^2, \ 9$

b. $1 + x, \ x^2, \ 2 + 2x + 3x^2$

c. $1 + x - 3x^2, \ 2 + 2x - 6x^2, \ 3 + 3x - 9x^2$

True-False Exercises

TF. In parts **(a)–(k)** determine whether the statement is true or false, and justify your answer.

a. The zero vector space has dimension zero.

b. There is a set of 17 linearly independent vectors in R^{17}.

c. There is a set of 11 vectors that span R^{17}.

d. Every linearly independent set of five vectors in R^5 is a basis for R^5.

e. Every set of five vectors that spans R^5 is a basis for R^5.

f. Every set of vectors that spans R^n contains a basis for R^n.

g. Every linearly independent set of vectors in R^n is contained in some basis for R^n.

h. There is a basis for M_{22} consisting of invertible matrices.

i. If A has size $n \times n$ and $I_n, A, A^2, \ldots, A^{n^2}$ are distinct matrices, then $\{I_n, A, A^2, \ldots, A^{n^2}\}$ is a linearly dependent set.

j. There are at least two distinct three-dimensional subspaces of P_2.

k. There are only three distinct two-dimensional subspaces of P_2.

Working with Technology

T1. Devise three different procedures for using your technology utility to determine the dimension of the subspace spanned by a set of vectors in R^n, and then use each of those procedures to determine the dimension of the subspace of R^5 spanned by the vectors

$$\mathbf{v}_1 = (2, 2, -1, 0, 1), \quad \mathbf{v}_2 = (-1, -1, 2, -3, 1),$$
$$\mathbf{v}_3 = (1, 1, -2, 0, -1), \quad \mathbf{v}_4 = (0, 0, 1, 1, 1)$$

T2. Find a basis for the row space of A by starting at the top and successively removing each row that is a linear combination of its predecessors.

$$A = \begin{bmatrix} 3.4 & 2.2 & 1.0 & -1.8 \\ 2.1 & 3.6 & 4.0 & -3.4 \\ 8.9 & 8.0 & 6.0 & 7.0 \\ 7.6 & 9.4 & 9.0 & -8.6 \\ 1.0 & 2.2 & 0.0 & 2.2 \end{bmatrix}$$

4.7 Change of Basis

A basis that is suitable for one problem may not be suitable for another, so it is a common process in the study of vector spaces to change from one basis to another. Because a basis is the vector space generalization of a coordinate system, changing bases is akin to changing coordinate axes in R^2 and R^3. In this section we will study problems related to changing bases.

Coordinate Maps

If $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is a basis for a finite-dimensional vector space V, and if

$$(\mathbf{v})_S = (c_1, c_2, \dots, c_n)$$

is the coordinate vector of $\mathbf{v}$ relative to S, then, as illustrated in Figure 4.5.6, the mapping

$$\mathbf{v} \to (\mathbf{v})_S \tag{1}$$

creates a connection (a one-to-one correspondence) between vectors in the *general* vector space V and vectors in the *Euclidean* vector space R^n. We call (1) the ***coordinate map relative to S*** from V to R^n. In this section we will find it convenient to express coordinate vectors in the matrix form

$$[\mathbf{v}]_S = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \tag{2}$$

where the square brackets emphasize the matrix notation (**Figure 4.7.1**).

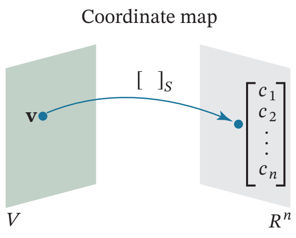


FIGURE 4.7.1

Change of Basis

There are many applications in which it is necessary to work with more than one coordinate system. In such cases it becomes important to know how the coordinates of a fixed vector relative to each coordinate system are related. This leads to the following problem.

The Change-of-Basis Problem

If $\mathbf{v}$ is a vector in a finite-dimensional vector space V, and if we change the basis for V from a basis B to a basis B', how are the coordinate vectors $[\mathbf{v}]_B$ and $[\mathbf{v}]_{B'}$ related?

Remark To solve this problem, it will be convenient to refer to the starting basis B as the "old basis" and the ending basis B' as the "new basis." Thus, our objective is to find a relationship between the old and new coordinates of a fixed vector $\mathbf{v}$ in V.

For simplicity, we will solve this problem for two-dimensional spaces. The solution for n-dimensional spaces is similar. Let

$$B = \{\mathbf{u}_1, \mathbf{u}_2\} \quad \text{and} \quad B' = \{\mathbf{u}'_1, \mathbf{u}'_2\}$$

be the old and new bases, respectively. Suppose that the coordinate vectors for the old basis vectors relative to the new basis are

$$[\mathbf{u}_1]_{B'} = \begin{bmatrix} a \\ b \end{bmatrix} \quad \text{and} \quad [\mathbf{u}_2]_{B'} = \begin{bmatrix} c \\ d \end{bmatrix} \tag{3}$$

That is,

$$\begin{aligned} \mathbf{u}_1 &= a\mathbf{u}'_1 + b\mathbf{u}'_2 \\ \mathbf{u}_2 &= c\mathbf{u}'_1 + d\mathbf{u}'_2 \end{aligned} \tag{4}$$

Now let $\mathbf{v}$ be any vector in V, and suppose that the old coordinate vector for $\mathbf{v}$ is

$$[\mathbf{v}]_B = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \tag{5}$$

so that

$$\mathbf{v} = k_1\mathbf{u}_1 + k_2\mathbf{u}_2 \tag{6}$$

In order to find the new coordinates of the vector $\mathbf{v}$ we must express $\mathbf{v}$ in terms of the new basis B'. To do this we will substitute (4) into (6), which yields

$$\mathbf{v} = k_1(a\mathbf{u}_1' + b\mathbf{u}_2') + k_2(c\mathbf{u}_1' + d\mathbf{u}_2')$$

or

$$\mathbf{v} = (k_1a + k_2c)\mathbf{u}_1' + (k_1b + k_2d)\mathbf{u}_2'$$

Thus, the new coordinate vector for $\mathbf{v}$ is

$$[\mathbf{v}]_{B'} = \begin{bmatrix} k_1a + k_2c \\ k_1b + k_2d \end{bmatrix}$$

which, by using (5), we can rewrite as

$$[\mathbf{v}]_{B'} = \begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} a & c \\ b & d \end{bmatrix}[\mathbf{v}]_B$$

This equation states that the new coordinate vector $[\mathbf{v}]_{B'}$ results when the old coordinate vector is multiplied on the left by the matrix

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

whose columns are the coordinate vectors of the old basis relative to the new basis [see (3)]. Thus, we are led to the following solution to the change-of-basis problem.

Solution to the Change-of-Basis Problem

If we change the basis for a vector space V from an old basis $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ to a new basis $B' = \{\mathbf{u}_1', \mathbf{u}_2', \ldots, \mathbf{u}_n'\}$, then for each vector $\mathbf{v}$ in V, the new coordinate vector $[\mathbf{v}]_{B'}$ is related to the old coordinate vector $[\mathbf{v}]_B$ by the equation

$$[\mathbf{v}]_{B'} = P[\mathbf{v}]_B \tag{7}$$

where the columns of P are the coordinate vectors of the old basis vectors relative to the new basis; that is

$$P = \Big[[\mathbf{u}_1]_{B'} \,|\, [\mathbf{u}_2]_{B'} \,|\, \cdots \,|\, [\mathbf{u}_n]_{B'}\Big] \tag{8}$$

Transition Matrices

The matrix P in Equations (7) and (8) is called the ***transition matrix from B to B′*** and will be denoted in this text as

$$P_{B\to B'} = \Big[[\mathbf{u}_1]_{B'} \,|\, [\mathbf{u}_2]_{B'} \,|\, \cdots \,|\, [\mathbf{u}_n]_{B'}\Big] \tag{9}$$

to emphasize that it changes coordinates relative to B into coordinates relative to B'. Analogously, the ***transition matrix from B′ to B*** will be denoted by

$$P_{B'\to B} = \Big[[\mathbf{u}_1']_B \,|\, [\mathbf{u}_2']_B \,|\, \cdots \,|\, [\mathbf{u}_n']_B\Big] \tag{10}$$

Remark In Formula (9) the old basis is B, and in Formula (10) the old basis is B'. Rather than memorizing these formulas, think about both in the following way.

The columns of the transition matrix from an old basis to a new basis are the coordinate vectors of the old basis relative to the new basis.

EXAMPLE 1 | Finding Transition Matrices

Consider the bases $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $B' = \{\mathbf{u}'_1, \mathbf{u}'_2\}$ for R^2, where

$$\mathbf{u}_1 = (1, 0), \quad \mathbf{u}_2 = (0, 1), \quad \mathbf{u}'_1 = (1, 1), \quad \mathbf{u}'_2 = (2, 1)$$

(*a*) Find the transition matrix $P_{B\to B'}$ from B to B'.

(*b*) Find the transition matrix $P_{B'\to B}$ from B' to B.

Solution (*a*) Here the old basis vectors are $\mathbf{u}_1$ and $\mathbf{u}_2$ and the new basis vectors are $\mathbf{u}'_1$ and $\mathbf{u}'_2$. We want to find the coordinate matrices of the old basis vectors relative to the new basis vectors. To do this, observe that

$$\begin{aligned} \mathbf{u}_1 &= -\mathbf{u}'_1 + \mathbf{u}'_2 \\ \mathbf{u}_2 &= 2\mathbf{u}'_1 - \mathbf{u}'_2 \end{aligned}$$

from which it follows that

$$[\mathbf{u}_1]_{B'} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \text{and} \quad [\mathbf{u}_2]_{B'} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

and hence that

$$P_{B\to B'} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}$$

Solution (*b*) Here the old basis vectors are $\mathbf{u}'_1$ and $\mathbf{u}'_2$ and the new basis vectors are $\mathbf{u}_1$ and $\mathbf{u}_2$. We want to find the coordinate matrices of the old basis vectors relative to the new basis vectors. To do this, observe that

$$\begin{aligned} \mathbf{u}'_1 &= \mathbf{u}_1 + \mathbf{u}_2 \\ \mathbf{u}'_2 &= 2\mathbf{u}_1 + \mathbf{u}_2 \end{aligned}$$

from which it follows that

$$[\mathbf{u}'_1]_{B} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad [\mathbf{u}'_2]_{B} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

and hence that

$$P_{B'\to B} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$$

Transforming Coordinates

Suppose now that B and B' are bases for a finite-dimensional vector space V. Since multiplication by $P_{B\to B'}$ maps coordinate vectors relative to the basis B into coordinate vectors relative to a basis B', and $P_{B'\to B}$ maps coordinate vectors relative to B' into coordinate vectors relative to B, it follows that for every vector $\mathbf{v}$ in V we have

$$\boxed{[\mathbf{v}]_{B'} = P_{B\to B'}[\mathbf{v}]_B} \tag{11}$$

$$\boxed{[\mathbf{v}]_{B} = P_{B'\to B}[\mathbf{v}]_{B'}} \tag{12}$$

EXAMPLE 2 | Change of Coordinates

Let B and B' be the bases in Example 1. Use an appropriate formula to find $[\mathbf{v}]_{B'}$ given that

$$[\mathbf{v}]_B = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$$

Solution To find $[\mathbf{v}]_{B'}$ we need to make the transition from B to B'. It follows from Formula (12) and part (*a*) of Example 1 that

$$[\mathbf{v}]_{B'} = P_{B\to B'}[\mathbf{v}]_B = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} -3 \\ 5 \end{bmatrix} = \begin{bmatrix} 13 \\ -8 \end{bmatrix}$$

Invertibility of Transition Matrices

If B and B' are bases for a finite-dimensional vector space V, then

$$(P_{B'\to B})(P_{B\to B'}) = P_{B\to B}$$

because multiplication by the product $(P_{B'\to B})(P_{B\to B'})$ first maps the B-coordinates of a vector into its B'-coordinates, and then maps those B'-coordinates back into the original B-coordinates. Since the net effect of the two operations is to leave each coordinate vector unchanged, we are led to conclude that $P_{B\to B}$ must be the identity matrix, that is,

$$(P_{B'\to B})(P_{B\to B'}) = I \tag{13}$$

For example, for the transition matrices obtained in Example 1 we have

$$(P_{B'\to B})(P_{B\to B'}) = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

It follows from (13) that $P_{B'\to B}$ is invertible and that its inverse is $P_{B\to B'}$. Thus, we have the following theorem.

Theorem 4.7.1

If P is the transition matrix from a basis B to a basis B' for a finite-dimensional vector space V, then P is invertible and P^{-1} is the transition matrix from B' to B.

An Efficient Method for Computing Transition Matrices between Bases for R^n

Our next objective is to develop an efficient procedure for computing transition matrices between bases for R^n. As illustrated in Example 1, the first step in computing a transition matrix is to express each new basis vector as a linear combination of the old basis vectors. For R^n this involves solving n linear systems of n equations in n unknowns, each of which has the same coefficient matrix (why?). An efficient way to do this is by the method illustrated in Example 2 of Section 1.6, which is as follows:

A Procedure for Computing Transition Matrices

Step 1. Form the partitioned matrix **[new basis | old basis]** in which the basis vectors are in column form.

Step 2. Use elementary row operations to reduce the matrix in Step 1 to reduced row echelon form.

Step 3. The resulting matrix will be **[I | transition matrix from old to new]** where I is an identity matrix.

Step 4. Extract the matrix on the right side of the matrix obtained in Step 3.

This procedure is captured in the diagram.

$$\boxed{[\text{new basis} \mid \text{old basis}] \xrightarrow{\text{row operations}} [I \mid \text{transition from old to new}]} \tag{14}$$

EXAMPLE 3 | Example 1 Revisited

In Example 1 we considered the bases $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $B' = \{\mathbf{u}'_1, \mathbf{u}'_2\}$ for R^2, where

$$\mathbf{u}_1 = (1, 0), \quad \mathbf{u}_2 = (0, 1), \quad \mathbf{u}'_1 = (1, 1), \quad \mathbf{u}'_2 = (2, 1)$$

(*a*) Use Formula (14) to find the transition matrix from B to B'.

(*b*) Use Formula (14) to find the transition matrix from B' to B.

Solution (*a*) Here B is the old basis and B' is the new basis, so

$$[\text{new basis} \mid \text{old basis}] = \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right]$$

By reducing this matrix, so the left side becomes the identity, we obtain (verify)

$$[I \mid \text{transition from old to new}] = \left[\begin{array}{cc|cc} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \end{array}\right]$$

so the transition matrix is

$$P_{B \to B'} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}$$

which agrees with the result in Example 1.

Solution (*b*) Here B' is the old basis and B is the new basis, so

$$[\text{new basis} \mid \text{old basis}] = \left[\begin{array}{cc|cc} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \end{array}\right]$$

Since the left side is already the identity matrix, no reduction is needed. We see by inspection that the transition matrix is

$$P_{B' \to B} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$$

which agrees with the result in Example 1.

Transition to the Standard Basis for R^n

Note that in part (*b*) of the last example the column vectors of the matrix that made the transition from the basis B to the standard basis turned out to be the vectors in B written in column form. This illustrates the following general result.

Theorem 4.7.2

Let $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be any basis for R^n and let $S = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ be the standard basis for R^n. If the vectors in these bases are written in column form, then

$$P_{B \to S} = [\mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_n] \tag{15}$$

It follows from this theorem that if

$$A = [\mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_n]$$

is any *invertible* $n \times n$ matrix, then A can be viewed as the transition matrix from the basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ for R^n to the standard basis for R^n. Thus, for example, the matrix

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

which was shown to be invertible in Example 4 of Section 1.5, is the transition matrix from the basis

$$\mathbf{u}_1 = (1, 2, 1), \quad \mathbf{u}_2 = (2, 5, 0), \quad \mathbf{u}_3 = (3, 3, 8)$$

to the basis

$$\mathbf{e}_1 = (1, 0, 0), \quad \mathbf{e}_2 = (0, 1, 0), \quad \mathbf{e}_3 = (0, 0, 1)$$

Exercise Set 4.7

1. Consider the bases $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $B' = \{\mathbf{u}'_1, \mathbf{u}'_2\}$ for R^2, where

$$\mathbf{u}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 4 \\ -1 \end{bmatrix}, \quad \mathbf{u}'_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad \mathbf{u}'_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

a. Find the transition matrix from B' to B.

b. Find the transition matrix from B to B'.

c. Compute the coordinate vector $[\mathbf{w}]_B$, where

$$\mathbf{w} = \begin{bmatrix} 3 \\ -5 \end{bmatrix}$$

and use (11) to compute $[\mathbf{w}]_{B'}$.

d. Check your work by computing $[\mathbf{w}]_{B'}$ directly.

2. Repeat the directions of Exercise 1 with the same vector $\mathbf{w}$ but with

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{u}'_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}'_2 = \begin{bmatrix} -3 \\ 4 \end{bmatrix}$$

3. Consider the bases $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and $B' = \{\mathbf{u}'_1, \mathbf{u}'_2, \mathbf{u}'_3\}$ for R^3, where

$$\mathbf{u}_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$\mathbf{u}'_1 = \begin{bmatrix} 3 \\ 1 \\ -5 \end{bmatrix}, \quad \mathbf{u}'_2 = \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix}, \quad \mathbf{u}'_3 = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$$

a. Find the transition matrix from B to B'.

b. Compute the coordinate vector $[\mathbf{w}]_B$, where

$$\mathbf{w} = \begin{bmatrix} -5 \\ 8 \\ -5 \end{bmatrix}$$

and use (11) to compute $[\mathbf{w}]_{B'}$.

c. Check your work by computing $[\mathbf{w}]_{B'}$ directly.

4. Repeat the directions of Exercise 3 with the same vector $\mathbf{w}$, but with

$$\mathbf{u}_1 = \begin{bmatrix} -3 \\ 0 \\ -3 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -3 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 1 \\ 6 \\ -1 \end{bmatrix}$$

$$\mathbf{u}'_1 = \begin{bmatrix} -6 \\ -6 \\ 0 \end{bmatrix}, \quad \mathbf{u}'_2 = \begin{bmatrix} -2 \\ -6 \\ 4 \end{bmatrix}, \quad \mathbf{u}'_3 = \begin{bmatrix} -2 \\ -3 \\ 7 \end{bmatrix}$$

5. Let V be the space spanned by $\mathbf{f}_1 = \sin x$ and $\mathbf{f}_2 = \cos x$.

a. Show that $\mathbf{g}_1 = 2\sin x + \cos x$ and $\mathbf{g}_2 = 3\cos x$ form a basis for V.

b. Find the transition matrix from $B' = \{\mathbf{g}_1, \mathbf{g}_2\}$ to $B = \{\mathbf{f}_1, \mathbf{f}_2\}$.

c. Find the transition matrix from B to B'.

d. Compute the coordinate vector $[\mathbf{h}]_B$, where $\mathbf{h} = 2\sin x - 5\cos x$, and use (11) to obtain $[\mathbf{h}]_{B'}$.

e. Check your work by computing $[\mathbf{h}]_{B'}$ directly.

6. Consider the bases $B = \{\mathbf{p}_1, \mathbf{p}_2\}$ and $B' = \{\mathbf{q}_1, \mathbf{q}_2\}$ for P_1, where

$$\mathbf{p}_1 = 6 + 3x, \quad \mathbf{p}_2 = 10 + 2x, \quad \mathbf{q}_1 = 2, \quad \mathbf{q}_2 = 3 + 2x$$

a. Find the transition matrix from B' to B.

b. Find the transition matrix from B to B'.

c. Compute the coordinate vector $[\mathbf{p}]_B$, where $\mathbf{p} = -4 + x$, and use (11) to compute $[\mathbf{p}]_{B'}$.

d. Check your work by computing $[\mathbf{p}]_{B'}$ directly.

7. Let $B_1 = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $B_2 = \{\mathbf{v}_1, \mathbf{v}_2\}$ be the bases for R^2 in which $\mathbf{u}_1 = (1, 2)$, $\mathbf{u}_2 = (2, 3)$, $\mathbf{v}_1 = (1, 3)$, and $\mathbf{v}_2 = (1, 4)$.

a. Use Formula (14) to find the transition matrix $P_{B_2 \to B_1}$.

b. Use Formula (14) to find the transition matrix $P_{B_1 \to B_2}$.

c. Confirm that the matrices $P_{B_2 \to B_1}$ and $P_{B_1 \to B_2}$ are inverses of one another.

d. Let $\mathbf{w} = (0, 1)$. Find $[\mathbf{w}]_{B_1}$ and then use the matrix $P_{B_1 \to B_2}$ to compute $[\mathbf{w}]_{B_2}$ from $[\mathbf{w}]_{B_1}$.

e. Let $\mathbf{w} = (2, 5)$. Find $[\mathbf{w}]_{B_2}$ and then use the matrix $P_{B_2 \to B_1}$ to compute $[\mathbf{w}]_{B_1}$ from $[\mathbf{w}]_{B_2}$.

8. Let S be the standard basis for R^2, and let $B = \{\mathbf{v}_1, \mathbf{v}_2\}$ be the basis in which $\mathbf{v}_1 = (2, 1)$ and $\mathbf{v}_2 = (-3, 4)$.

a. Find the transition matrix $P_{B \to S}$ by inspection.

b. Use Formula (14) to find the transition matrix $P_{S \to B}$.

c. Confirm that $P_{B \to S}$ and $P_{S \to B}$ are inverses of one another.

d. Let $\mathbf{w} = (5, -3)$. Find $[\mathbf{w}]_B$ and then use Formula (12) to compute $[\mathbf{w}]_S$.

e. Let $\mathbf{w} = (3, -5)$. Find $[\mathbf{w}]_S$ and then use Formula (11) to compute $[\mathbf{w}]_B$.

9. Let S be the standard basis for R^3, and let $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ be the basis in which $\mathbf{v}_1 = (1, 2, 1)$, $\mathbf{v}_2 = (2, 5, 0)$, and $\mathbf{v}_3 = (3, 3, 8)$.

a. Find the transition matrix $P_{B \to S}$ by inspection.

b. Use Formula (14) to find the transition matrix $P_{S \to B}$.

c. Confirm that $P_{B \to S}$ and $P_{S \to B}$ are inverses of one another.

d. Let $\mathbf{w} = (5, -3, 1)$. Find $[\mathbf{w}]_B$ and then use Formula (12) to compute $[\mathbf{w}]_S$.

e. Let $\mathbf{w} = (3, -5, 0)$. Find $[\mathbf{w}]_S$ and then use Formula (11) to compute $[\mathbf{w}]_B$.

10. Let $S = \{\mathbf{e}_1, \mathbf{e}_2\}$ be the standard basis for the vector space R^2, and let $B = \{\mathbf{v}_1, \mathbf{v}_2\}$ be the basis that results when the vectors in S are reflected about the line $y = x$.

a. Find the transition matrix $P_{B \to S}$.

b. Let $P = P_{B \to S}$ and show that $P^T = P_{S \to B}$.

11. Let $S = \{\mathbf{e}_1, \mathbf{e}_2\}$ be the standard basis for the vector space R^2, and let $B = \{\mathbf{v}_1, \mathbf{v}_2\}$ be the basis that results when the vectors in S are reflected about the line that makes an angle θ with the positive x-axis.

a. Find the transition matrix $P_{B \to S}$.

b. Let $P = P_{B \to S}$ and show that $P^T = P_{S \to B}$.

12. If B_1, B_2, and B_3 are bases for R^2, and if

$$P_{B_1 \to B_2} = \begin{bmatrix} 3 & 1 \\ 5 & 2 \end{bmatrix} \quad \text{and} \quad P_{B_2 \to B_3} = \begin{bmatrix} 7 & 2 \\ 4 & -1 \end{bmatrix}$$

then $P_{B_3 \to B_1} =$ ________.

13. If P is the transition matrix from a basis B' to a basis B, and Q is the transition matrix from B to a basis C, what is the transition matrix from B' to C? What is the transition matrix from C to B'?

14. To write the coordinate vector for a vector, it is necessary to specify an order for the vectors in the basis. If P is the transition matrix from a basis B' to a basis B, what is the effect on P if we reverse the order of vectors in B from $\mathbf{v}_1, \ldots, \mathbf{v}_n$ to $\mathbf{v}_n, \ldots, \mathbf{v}_1$? What is the effect on P if we reverse the order of vectors in both B' and B?

15. Consider the matrix

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

a. P is the transition matrix from what basis B to the standard basis $S = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ for R^3?

b. P is the transition matrix from the standard basis $S = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ to what basis B for R^3?

16. The matrix

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

is the transition matrix from what basis B to the basis $\{(1, 1, 1), (1, 1, 0), (1, 0, 0)\}$ for R^3?

17. Let $S = \{\mathbf{e}_1, \mathbf{e}_2\}$ be the standard basis for R^2, and let $B = \{\mathbf{v}_1, \mathbf{v}_2\}$ be the basis that results when the linear transformation defined by

$$T(x_1, x_2) = (2x_1 + 3x_2, 5x_1 - x_2)$$

is applied to each vector in S. Find the transition matrix $P_{B \to S}$.

18. Let $S = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be the standard basis for the vector space R^3, and let $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ be the basis that results when the linear transformation defined by

$$T(x_1, x_2, x_3) = (x_1 + x_2, 2x_1 - x_2 + 4x_3, x_2 + 3x_3)$$

is applied to each vector in S. Find the transition matrix $P_{B \to S}$.

19. If $[\mathbf{w}]_B = \mathbf{w}$ holds for all vectors $\mathbf{w}$ in R^n, what can you say about the basis B?

Working with Proofs

20. Let B be a basis for R^n. Prove that the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ span R^n if and only if the vectors $[\mathbf{v}_1]_B, [\mathbf{v}_2]_B, \ldots, [\mathbf{v}_k]_B$ span R^n.

21. Let B be a basis for R^n. Prove that the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ form a linearly independent set in R^n if and only if the vectors $[\mathbf{v}_1]_B, [\mathbf{v}_2]_B, \ldots, [\mathbf{v}_k]_B$ form a linearly independent set in R^n.

True-False Exercises

TF. In parts **(a)–(f)** determine whether the statement is true or false, and justify your answer.

a. If B_1 and B_2 are bases for a vector space V, then there exists a transition matrix from B_1 to B_2.

b. Transition matrices are invertible.

c. If B is a basis for a vector space R^n, then $P_{B \to B}$ is the identity matrix.

d. If $P_{B_1 \to B_2}$ is a diagonal matrix, then each vector in B_2 is a scalar multiple of some vector in B_1.

e. If each vector in B_2 is a scalar multiple of some vector in B_1, then $P_{B_1 \to B_2}$ is a diagonal matrix.

f. If A is a square matrix, then $A = P_{B_1 \to B_2}$ for some bases B_1 and B_2 for R^n.

Working with Technology

T1. Let

$$P = \begin{bmatrix} 5 & 8 & 6 & -13 \\ 3 & -1 & 0 & -9 \\ 0 & 1 & -1 & 0 \\ 2 & 4 & 3 & -5 \end{bmatrix}$$

and

$$\mathbf{v}_1 = (2, 4, 3, -5), \quad \mathbf{v}_2 = (0, 1, -1, 0),$$
$$\mathbf{v}_3 = (3, -1, 0, -9), \quad \mathbf{v}_4 = (5, 8, 6, -13)$$

Find a basis $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ for R^4 for which P is the transition matrix from B to $B' = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$.

T2. Given that the matrix for a linear transformation $T: R^4 \to R^4$ relative to the standard basis $B = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ for R^4 is

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

find the matrix for T relative to the basis

$$B' = \{\mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3, \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3 + \mathbf{e}_4\}$$

4.8 Row Space, Column Space, and Null Space

In this section we will study some important vector spaces that are associated with matrices. Our work here will provide us with a deeper understanding of the relationships between the solutions of a linear system and properties of its coefficient matrix.

Matrix Spaces

Recall that vectors can be written in comma-delimited form or in matrix form as either row vectors or column vectors. In this section we will use the latter two.

Definition 1

For an $m \times n$ matrix

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

the vectors

$$\begin{aligned} \mathbf{r}_1 &= [a_{11} \quad a_{12} \quad \cdots \quad a_{1n}] \\ \mathbf{r}_2 &= [a_{21} \quad a_{22} \quad \cdots \quad a_{2n}] \\ &\ \,\vdots \qquad\qquad \vdots \\ \mathbf{r}_m &= [a_{m1} \quad a_{m2} \quad \cdots \quad a_{mn}] \end{aligned}$$

in R^n formed from the rows of A are called the ***row vectors*** of A, and the vectors

$$\mathbf{c}_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}, \ldots, \quad \mathbf{c}_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

in R^m formed from the columns of A are called the ***column vectors*** of A.

EXAMPLE 1 | Row and Column Vectors of a 2 × 3 Matrix

Let

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

The row vectors of A are

$$\mathbf{r}_1 = [2 \quad 1 \quad 0] \quad \text{and} \quad \mathbf{r}_2 = [3 \quad -1 \quad 4]$$

and the column vectors of A are

$$\mathbf{c}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \text{and} \quad \mathbf{c}_3 = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$$

The following definition defines three important vector spaces associated with a matrix.

Definition 2

If A is an $m \times n$ matrix, then the subspace of R^n spanned by the row vectors of A is denoted by row(A) and is called the ***row space*** of A, and the subspace of R^m spanned by the column vectors of A is denoted by col(A) and is called the ***column space*** of A. The solution space of the homogeneous system of equations $A\mathbf{x} = \mathbf{0}$, which is a subspace of R^n, is denoted by null(A) and is called the ***null space*** of A.

Throughout this section and the next we will consider with two general questions:

Question 1. What relationships exist among the solutions of a linear system $A\mathbf{x} = \mathbf{b}$ and the row space, column space, and null space of the coefficient matrix A?

Question 2. What relationships exist among the row space, column space, and null space of a matrix?

Starting with the first question, suppose that

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

It follows from Formula (10) of Section 1.3 that if $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ denote the column vectors of A, then the product $A\mathbf{x}$ can be expressed as a linear combination of these vectors with coefficients from $\mathbf{x}$; that is,

$$A\mathbf{x} = x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n \tag{1}$$

Thus, a linear system, $A\mathbf{x} = \mathbf{b}$, of m equations in n unknowns can be written as

$$x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n = \mathbf{b} \tag{2}$$

from which we conclude that $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is expressible as a linear combination of the column vectors of A. This yields the following theorem.

Theorem 4.8.1

A system of linear equations $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is in the column space of A.

EXAMPLE 2 | A Vector b in the Column Space of A

Let $A\mathbf{x} = \mathbf{b}$ be the linear system

$$\begin{bmatrix} -1 & 3 & 2 \\ 1 & 2 & -3 \\ 2 & 1 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -9 \\ -3 \end{bmatrix}$$

Show that $\mathbf{b}$ is in the column space of A by expressing it as a linear combination of the column vectors of A.

Solution Solving the system by Gaussian elimination yields (verify)

$$x_1 = 2, \quad x_2 = -1, \quad x_3 = 3$$

It follows from this and Formula (2) that

$$2\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} + 3\begin{bmatrix} 2 \\ -3 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ -9 \\ -3 \end{bmatrix}$$

The Relationship Between $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{x} = \mathbf{b}$

In this subsection we will explore the relationship between the solutions of a homogeneous linear system $A\mathbf{x} = \mathbf{0}$ and the solutions (if any) of the nonhomogeneous linear system $A\mathbf{x} = \mathbf{b}$ with the same coefficient matrix. These are called ***corresponding linear systems.*** By way of example, we will consider the following linear systems that we first discussed in Examples 5 and 6 of Section 1.2 and then again in Example 3 of Section 4.6.

$$\begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 \\ 0 & 0 & 5 & 10 & 0 & 15 \\ 2 & 6 & 0 & 8 & 4 & 18 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 \\ 0 & 0 & 5 & 10 & 0 & 15 \\ 2 & 6 & 0 & 8 & 4 & 18 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 5 \\ 6 \end{bmatrix}$$

In Section 1.2 we found the general solutions of these systems to be

homogeneous $\longrightarrow x_1 = -3r - 4s - 2t, \quad x_2 = r, \quad x_3 = -2s, \quad x_4 = s, \quad x_5 = t, \quad x_6 = 0$

nonhomogeneous $\longrightarrow x_1 = -3r - 4s - 2t, \quad x_2 = r, \quad x_3 = -2s, \quad x_4 = s, \quad x_5 = t, \quad x_6 = \frac{1}{3}$

which we can express in column-vector form as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} -3r - 4s - 2t \\ r \\ -2s \\ s \\ t \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} -3r - 4s - 2t \\ r \\ -2s \\ s \\ t \\ \frac{1}{3} \end{bmatrix}$$

By splitting the entries on the right apart and collecting terms with like parameters we can rewrite these general solutions as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = r\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} -4 \\ 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tag{3}$$

Homogeneous Case

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = r\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} -4 \\ 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{3} \end{bmatrix} \tag{4}$$

Nonhomogeneous Case

In Example 3 of Section 4.6 we observed that the three vectors on the right side of (3) are linearly independent and therefore form a basis for the solution space of the homogeneous system. Thus, as illustrated in (5), the general solution $\mathbf{x}$ of the nonhomogeneous system can be divided into two parts, a basis $\mathbf{x}_h$ for the null space of the homogeneous system and a term $\mathbf{x}_0$ that is a solution of the nonhomogeneous system (in this case, the solution resulting from setting the parameters to zero).

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}}_{\mathbf{x}} = \begin{bmatrix} -3r-4s-2t \\ r \\ -2s \\ s \\ t \\ \frac{1}{3} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{3} \end{bmatrix}}_{\mathbf{x}_0} + \underbrace{r\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} -4 \\ 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{x}_h} \tag{5}$$

This example illustrates the following general theorem.

Theorem 4.8.2

If $\mathbf{x}_0$ is any solution of a consistent linear system $A\mathbf{x} = \mathbf{b}$, and if $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a basis for the null space of A, then every solution of $A\mathbf{x} = \mathbf{b}$ can be expressed in the form

$$\mathbf{x} = \mathbf{x}_0 + c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k \tag{6}$$

Conversely, for all choices of scalars $c_1, c_2, \ldots, c_k$, the vector $\mathbf{x}$ in this formula is a solution of $A\mathbf{x} = \mathbf{b}$.

Proof Let $\mathbf{x}_0$ be any solution of $A\mathbf{x} = \mathbf{b}$, let W denote the null space of $A\mathbf{x} = \mathbf{0}$, and let $\mathbf{x}_0 + W$ be the set of all vectors that result by adding $\mathbf{x}_0$ to each vector in W. Thus, the vectors in $\mathbf{x}_0 + W$ are those that are expressible in the form

$$\mathbf{x} = \mathbf{x}_0 + c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

We must show that if $\mathbf{x}$ is a vector in $\mathbf{x}_0 + W$, then $\mathbf{x}$ is a solution of $A\mathbf{x} = \mathbf{b}$, and conversely that every solution of $A\mathbf{x} = \mathbf{b}$ is in the set $\mathbf{x}_0 + W$.

Assume first that $\mathbf{x}$ is a vector in $\mathbf{x}_0 + W$. This implies that $\mathbf{x}$ is expressible in the form $\mathbf{x} = \mathbf{x}_0 + \mathbf{w}$, where $A\mathbf{x}_0 = \mathbf{b}$ and $A\mathbf{w} = \mathbf{0}$. Thus,

$$A\mathbf{x} = A(\mathbf{x}_0 + \mathbf{w}) = A\mathbf{x}_0 + A\mathbf{w} = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

which shows that $\mathbf{x}$ is a solution of $A\mathbf{x} = \mathbf{b}$.

Conversely, let $\mathbf{x}$ be any solution of $A\mathbf{x} = \mathbf{b}$. To show that $\mathbf{x}$ is in the set $\mathbf{x}_0 + W$ we must show that $\mathbf{x}$ is expressible in the form

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{w} \tag{7}$$

where $\mathbf{w}$ is in W (i.e., $A\mathbf{w} = \mathbf{0}$). We can do this by taking $\mathbf{w} = \mathbf{x} - \mathbf{x}_0$. This vector obviously satisfies (7), and it is in W since

$$A\mathbf{w} = A(\mathbf{x} - \mathbf{x}_0) = A\mathbf{x} - A\mathbf{x}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0} \quad \blacksquare$$

The vector $\mathbf{x}_0$ in Formula (6) is called a ***particular solution of*** $A\mathbf{x} = \mathbf{b}$, and the remaining part of the formula is called the ***general solution of*** $A\mathbf{x} = \mathbf{0}$. With this terminology Theorem 4.8.2 can be rephrased as:

> *The general solution of a consistent linear system can be expressed as the sum of a particular solution of that system and the general solution of the corresponding homogeneous system.*

Geometrically, the solution set of $A\mathbf{x} = \mathbf{b}$ can be viewed as the translation by $\mathbf{x}_0$ of the solution space of $A\mathbf{x} = \mathbf{0}$ (**Figure 4.8.1**).

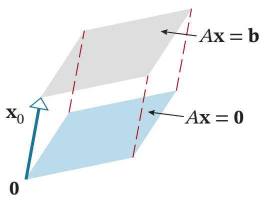


FIGURE 4.8.1 The solution space of $A\mathbf{x} = \mathbf{b}$ is a translation of the solution space of $A\mathbf{x} = \mathbf{0}$.

Bases for Row Spaces, Column Spaces, and Null Spaces

In this subsection we will focus on the second problem posed earlier in this section, finding relationships between the row space, column space, and null space of a matrix. We begin with the following theorem.

Theorem 4.8.3

(*a*) *Row equivalent matrices have the same row space.*

(*b*) *Row equivalent matrices have the same null space.*

Proof (a) If A and B are row equivalent then each can be obtained from the other by elementary row operations. As these operations involve only scalar multiplication (multiply a row by a scalar) and linear combinations (add a scalar multiple of one row to another), it follows that the row space of each is a subspace of the other, so the two row spaces must be the same.

Proof (b) If A and B are row equivalent then each can be obtained from the other by elementary row operations. But elementary row operations do not change the solution set of a linear system, so the solution sets of $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ must be the same. That is, A and B have the same null space. $\blacksquare$

Theorem 4.8.3 might tempt you into *incorrectly* believing that elementary row operations do not change the column space of a matrix. To see why this is *not* true, compare the matrices

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

The matrix B can be obtained from A by adding -2 times the first row to the second. However, this operation has changed the column space of A, since that column space consists of all scalar multiples of

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

whereas the column space of B consists of all scalar multiples of

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

and the two are different spaces.

The following theorem makes it possible to find bases for the row and column spaces of a matrix in row echelon form by inspection.

Theorem 4.8.4

If a matrix R is in row echelon form, then the row vectors with the leading 1's (the nonzero row vectors) form a basis for the row space of R, and the column vectors with the leading 1's of the row vectors form a basis for the column space of R.

The proof essentially involves an analysis of the positions of the 0's and 1's of R. We omit the details.

EXAMPLE 3 | Bases for the Row and Column Spaces of a Matrix in Row Echelon Form

Find bases for the row and column spaces of the matrix

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

Solution Since the matrix R is in row echelon form, it follows from Theorem 4.8.4 that the vectors

$$\begin{aligned} \mathbf{r}_1 &= [1 \quad -2 \quad 5 \quad 0 \quad 3] \\ \mathbf{r}_2 &= [0 \quad 1 \quad 3 \quad 0 \quad 0] \\ \mathbf{r}_3 &= [0 \quad 0 \quad 0 \quad 1 \quad 0] \end{aligned}$$

form a basis for the row space of R, and the vectors

$$\mathbf{c}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{c}_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

form a basis for the column space of R.

Theorem 4.8.3(a) and Theorem 4.8.4 in combination make it possible to find a basis for the row space of a matrix A by reducing it to a row echelon form R.

EXAMPLE 4 | Basis for a Row Space by Row Reduction

Find a basis for the row space of the matrix

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

Solution Since elementary row operations do not change the row space of a matrix, we can find a basis for the row space of A by finding a basis for the row space of any row echelon form of A. Reducing A to row echelon form, we obtain (verify)

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

By Theorem 4.8.4, the nonzero row vectors of R form a basis for the row space of R and hence form a basis for the row space of A. These basis vectors are

$$\begin{aligned} \mathbf{r}_1 &= [1 \quad -3 \quad 4 \quad -2 \quad 5 \quad 4] \\ \mathbf{r}_2 &= [0 \quad 0 \quad 1 \quad 3 \quad -2 \quad -6] \\ \mathbf{r}_3 &= [0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 5] \end{aligned}$$

Bases Formed from Row and Column Vectors of a Matrix

If a matrix A is reduced to a row echelon form R, we know how to find a basis for the row space and column space of R (Example 3). Moreover, we also know that the basis obtained for the row space of R is a basis for the row space of A (Example 4). What is *not* true, however, is that the basis obtained for the column space of R is also a basis for the column space of A, the problem being that elementary row operations can change column spaces. However, the good news is that *elementary row operations do not change dependency relationships between column vectors.* To make this precise, suppose that $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ are linearly dependent column vectors of A, so there are scalars $c_1, c_2, \ldots, c_k$ that are not all zero for which

$$c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_k\mathbf{w}_k = \mathbf{0} \tag{8}$$

If we perform an elementary row operation on A, then these vectors will be changed into new column vectors $\mathbf{w}'_1, \mathbf{w}'_2, \ldots, \mathbf{w}'_k$. At first glance it would seem possible that the transformed vectors might be linearly independent. However, this is not so, since it can be proved that these new column vectors are linearly dependent and, in fact, related by an equation

$$c_1\mathbf{w}'_1 + c_2\mathbf{w}'_2 + \cdots + c_k\mathbf{w}'_k = \mathbf{0}$$

that has exactly the same coefficients as (8). It can also be proved that elementary row operations do not alter the linear independence of a set of column vectors. All of these results are summarized in the following theorem.

Theorem 4.8.5

If A and B are row equivalent matrices, then:

(*a*) A given set of column vectors of A is linearly independent if and only if the corresponding column vectors of B are linearly independent.

(*b*) A given set of column vectors of A forms a basis for the column space of A if and only if the corresponding column vectors of B form a basis for the column space of B.

It follows from Theorem 4.8.5(*b*) that even though an elementary row operation can change the column space, it does not change the *dimension* of the column space.

EXAMPLE 5 | Basis from the Columns of A

Find a basis for the column space of the matrix

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

that consists of column vectors of A.

Solution We observed in Example 4 that the matrix

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

is a row echelon form of A. Keeping in mind that A and R can have different column spaces, we cannot find a basis for the column space of A directly from the column vectors of R. However, it follows from Theorem 4.8.5(b) that if we can find a set of column vectors of R that forms a basis for the column space of R, then the *corresponding* column vectors of A will form a basis for the column space of A.

Since the first, third, and fifth columns of R contain the leading 1's of the row vectors, the vectors

$$\mathbf{c}_1' = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{c}_3' = \begin{bmatrix} 4 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{c}_5' = \begin{bmatrix} 5 \\ -2 \\ 1 \\ 0 \end{bmatrix}$$

form a basis for the column space of R. Thus, the corresponding column vectors of A, which are

$$\mathbf{c}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{c}_3 = \begin{bmatrix} 4 \\ 9 \\ 9 \\ -4 \end{bmatrix}, \quad \mathbf{c}_5 = \begin{bmatrix} 5 \\ 8 \\ 9 \\ -5 \end{bmatrix}$$

form a basis for the column space of A.

In Example 4, we found a basis for the row space of a matrix by reducing that matrix to row echelon form. However, the basis vectors produced by that method were not all row vectors of the original matrix. The following adaptation of the technique used in Example 5 shows how to find a basis for the row space of a matrix that consists entirely of row vectors of that matrix.

EXAMPLE 6 | Basis from the Rows of A

Find a basis for the row space of

$$A = \begin{bmatrix} 1 & -2 & 0 & 0 & 3 \\ 2 & -5 & -3 & -2 & 6 \\ 0 & 5 & 15 & 10 & 0 \\ 2 & 6 & 18 & 8 & 6 \end{bmatrix}$$

consisting entirely of row vectors from A.

Solution We will transpose A, thereby converting the row space of A into the column space of A^T; then we will use the method of Example 5 to find a basis for the column space of A^T; and then we will transpose again to convert column vectors back to row vectors.

Transposing A yields

$$A^T = \begin{bmatrix} 1 & 2 & 0 & 2 \\ -2 & -5 & 5 & 6 \\ 0 & -3 & 15 & 18 \\ 0 & -2 & 10 & 8 \\ 3 & 6 & 0 & 6 \end{bmatrix}$$

and then reducing this matrix to row echelon form we obtain

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

The first, second, and fourth columns contain the leading 1's, so the corresponding column vectors in A^T form a basis for the column space of A^T; these are

$$\mathbf{c}_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 0 \\ 3 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} 2 \\ -5 \\ -3 \\ -2 \\ 6 \end{bmatrix}, \quad \text{and} \quad \mathbf{c}_4 = \begin{bmatrix} 2 \\ 6 \\ 18 \\ 8 \\ 6 \end{bmatrix}$$

Transposing again and adjusting the notation appropriately yields the basis vectors

$$\mathbf{r}_1 = [1 \quad -2 \quad 0 \quad 0 \quad 3], \qquad \mathbf{r}_2 = [2 \quad -5 \quad -3 \quad -2 \quad 6],$$
$$\mathbf{r}_4 = [2 \quad 6 \quad 18 \quad 8 \quad 6]$$

for the row space of A.

Up to now we have focused on methods for finding bases associated with matrices. Those methods can readily be adapted to the more general problem of finding a basis for the subspace spanned by a set of vectors in R^n.

EXAMPLE 7 | Basis for the Space Spanned by a Set of Vectors

The following vectors span a subspace of R^4. Find a subset of these vectors that forms a basis of this subspace.

$$\mathbf{v}_1 = (1, 2, 2, -1), \qquad \mathbf{v}_2 = (-3, -6, -6, 3),$$
$$\mathbf{v}_3 = (4, 9, 9, -4), \qquad \mathbf{v}_4 = (-2, -1, -1, 2),$$
$$\mathbf{v}_5 = (5, 8, 9, -5), \qquad \mathbf{v}_6 = (4, 2, 7, -4)$$

Solution If we rewrite these vectors in column form and construct the matrix that has those vectors as its successive columns, then we obtain the matrix A in Example 7 (verify). Thus,

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\} = \text{col}(A)$$

Proceeding as in that example (and adjusting the notation appropriately), we see that the vectors $\mathbf{v}_1$, $\mathbf{v}_3$, and $\mathbf{v}_5$ form a basis for

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}$$

Next we will give an example that adapts the method of Example 5 to solve the following general problem in R^n:

Problem

Given a set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in R^n, find a subset of these vectors that forms a basis for span(S), and express each vector that is not in that basis as a linear combination of the basis vectors.

EXAMPLE 8 | Basis and Linear Combinations

(*a*) Find a subset of the vectors

$$\mathbf{v}_1 = (1, -2, 0, 3), \quad \mathbf{v}_2 = (2, -5, -3, 6),$$
$$\mathbf{v}_3 = (0, 1, 3, 0), \quad \mathbf{v}_4 = (2, -1, 4, -7), \quad \mathbf{v}_5 = (5, -8, 1, 2)$$

that forms a basis for the subspace of R^4 spanned by these vectors.

(*b*) Express each vector not in the basis as a linear combination of the basis vectors.

Solution (*a*) We begin by constructing a matrix that has $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_5$ as its column vectors:

$$\begin{array}{c} \begin{bmatrix} 1 & 2 & 0 & 2 & 5 \\ -2 & -5 & 1 & -1 & -8 \\ 0 & -3 & 3 & 4 & 1 \\ 3 & 6 & 0 & -7 & 2 \end{bmatrix} \\ \begin{array}{ccccc} \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 & \mathbf{v}_5 \end{array} \end{array} \tag{9}$$

The first part of our problem can be solved by finding a basis for the column space of this matrix. Reducing the matrix to *reduced* row echelon form and denoting the column vectors of the resulting matrix by $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, $\mathbf{w}_4$, and $\mathbf{w}_5$ yields

$$\begin{array}{c} \begin{bmatrix} 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \begin{array}{ccccc} \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ \mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3 & \mathbf{w}_4 & \mathbf{w}_5 \end{array} \end{array} \tag{10}$$

The leading 1's occur in columns 1, 2, and 4, so by Theorem 4.8.4,

$$\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_4\}$$

is a basis for the column space of (6), and consequently,

$$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$$

is a basis for the column space of (9).

Solution (*b*) We will start by expressing $\mathbf{w}_3$ and $\mathbf{w}_5$ as linear combinations of the basis vectors $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_4$. The simplest way of doing this is to express $\mathbf{w}_3$ and $\mathbf{w}_5$ in terms of basis vectors with numerically smaller subscripts. Accordingly, we will express $\mathbf{w}_3$ as a linear combination of $\mathbf{w}_1$ and $\mathbf{w}_2$, and we will express $\mathbf{w}_5$ as a linear combination of the vectors $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_4$. By inspection of (10), these linear combinations are

$$\mathbf{w}_3 = 2\mathbf{w}_1 - \mathbf{w}_2$$
$$\mathbf{w}_5 = \mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_4$$

We call these the ***dependency equations***. The corresponding relationships in (9) are

$$\mathbf{v}_3 = 2\mathbf{v}_1 - \mathbf{v}_2$$
$$\mathbf{v}_5 = \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_4$$

Had we only been interested in part (*a*) of this example, it would have sufficed to reduce the matrix to row echelon form. It is for part (*b*) that the reduced row echelon form is most useful.

The following is a summary of the steps that we followed in our last example to solve the problem posed above.

Basis for the Space Spanned by a Set of Vectors

Step 1. Form the matrix A whose columns are the vectors in the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$.

Step 2. Reduce the matrix A to reduced row echelon form R.

Step 3. Denote the column vectors of R by $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$.

Step 4. Identify the columns of R that contain the leading 1's. The corresponding column vectors of A form a basis for span(S).

This completes the first part of the problem.

Step 5. Obtain a set of dependency equations for the column vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ of R by successively expressing each $\mathbf{w}_i$ that does not contain a leading 1 of R as a linear combination of predecessors that do.

Step 6. In each dependency equation obtained in Step 5, replace the vector $\mathbf{w}_i$ by the vector $\mathbf{v}_i$ for $i = 1, 2, \ldots, k$.

This completes the second part of the problem.

Exercise Set 4.8

In Exercises 1–2, express the product $A\mathbf{x}$ as a linear combination of the column vectors of A.

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

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

*In Exercises 3–4, determine whether **b** is in the column space of A, and if so, express **b** as a linear combination of the column vectors of A.*

3. a. $A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 3 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$

b. $A = \begin{bmatrix} 1 & -1 & 1 \\ 9 & 3 & 1 \\ 1 & 1 & 1 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 5 \\ 1 \\ -1 \end{bmatrix}$

4. a. $A = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$

b. $A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 2 & 1 \\ 1 & 2 & 1 & 3 \\ 0 & 1 & 2 & 2 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 4 \\ 3 \\ 5 \\ 7 \end{bmatrix}$

5. Suppose that $x_1 = 3, x_2 = 0, x_3 = -1, x_4 = 5$ is a solution of a nonhomogeneous linear system $A\mathbf{x} = \mathbf{b}$ and that the solution set of the homogeneous system $A\mathbf{x} = \mathbf{0}$ is given by the formulas

$$x_1 = 5r - 2s, \quad x_2 = s, \quad x_3 = s + t, \quad x_4 = t$$

a. Find a vector form of the general solution of $A\mathbf{x} = \mathbf{0}$.

b. Find a vector form of the general solution of $A\mathbf{x} = \mathbf{b}$.

6. Suppose that $x_1 = -1, x_2 = 2, x_3 = 4, x_4 = -3$ is a solution of a nonhomogeneous linear system $A\mathbf{x} = \mathbf{b}$ and that the solution set of the homogeneous system $A\mathbf{x} = \mathbf{0}$ is given by the formulas

$$x_1 = -3r + 4s, \quad x_2 = r - s, \quad x_3 = r, \quad x_4 = s$$

a. Find a vector form of the general solution of $A\mathbf{x} = \mathbf{0}$.

b. Find a vector form of the general solution of $A\mathbf{x} = \mathbf{b}$.

In Exercises 7–8, find the vector form of the general solution of the linear system $A\mathbf{x} = \mathbf{b}$, and then use that result to find the vector form of the general solution of $A\mathbf{x} = \mathbf{0}$.

7. a. $\begin{aligned} x_1 - 3x_2 &= 1 \\ 2x_1 - 6x_2 &= 2 \end{aligned}$ **b.** $\begin{aligned} x_1 + x_2 + 2x_3 &= 5 \\ x_1 + x_3 &= -2 \\ 2x_1 + x_2 + 3x_3 &= 3 \end{aligned}$

8. a. $\begin{aligned} x_1 - 2x_2 + x_3 + 2x_4 &= -1 \\ 2x_1 - 4x_2 + 2x_3 + 4x_4 &= -2 \\ -x_1 + 2x_2 - x_3 - 2x_4 &= 1 \\ 3x_1 - 6x_2 + 3x_3 + 6x_4 &= -3 \end{aligned}$

b.
$$\begin{aligned} x_1 + 2x_2 - 3x_3 + x_4 &= 4 \\ -2x_1 + x_2 + 2x_3 + x_4 &= -1 \\ -x_1 + 3x_2 - x_3 + 2x_4 &= 3 \\ 4x_1 - 7x_2 \qquad - 5x_4 &= -5 \end{aligned}$$

In Exercises 9–10, find bases for the null space and row space of A.

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

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

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

In Exercises 11–12, a matrix in row echelon form is given. By inspection, find a basis for the row space and for the column space of that matrix.

11. **a.** $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & -3 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

12. **a.** $\begin{bmatrix} 1 & 2 & 4 & 5 \\ 0 & 1 & -3 & 0 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 2 & -1 & 5 \\ 0 & 1 & 4 & 3 \\ 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

13. **a.** Use the methods of Examples 6 and 7 to find bases for the row space and column space of the matrix

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

b. Use the method of Example 9 to find a basis for the row space of A that consists entirely of row vectors of A.

In Exercises 14–15, find a basis for the subspace of R^4 that is spanned by the given vectors.

14. $(1, 1, -4, -3),\ (2, 0, 2, -2),\ (2, -1, 3, 2)$

15. $(1, 1, 0, 0),\ (0, 0, 1, 1),\ (-2, 0, 2, 2),\ (0, -3, 0, 3)$

In Exercises 16–17, find a subset of the given vectors that forms a basis for the space spanned by those vectors, and then express each vector that is not in the basis as a linear combination of the basis vectors.

16. $\mathbf{v}_1 = (1, 0, 1, 1),\ \mathbf{v}_2 = (-3, 3, 7, 1),$
$\mathbf{v}_3 = (-1, 3, 9, 3),\ \mathbf{v}_4 = (-5, 3, 5, -1)$

17. $\mathbf{v}_1 = (1, -1, 5, 2),\ \mathbf{v}_2 = (-2, 3, 1, 0),$
$\mathbf{v}_3 = (4, -5, 9, 4),\ \mathbf{v}_4 = (0, 4, 2, -3),$
$\mathbf{v}_5 = (-7, 18, 2, -8)$

In Exercises 18–19, find a basis for the row space of A that consists entirely of row vectors of A.

18. The matrix in Exercise 10(a).

19. The matrix in Exercise 10(b).

20. Construct a matrix whose null space consists of all linear combinations of the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 3 \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 0 \\ -2 \\ 4 \end{bmatrix}$$

21. In each part, let $A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & -1 & 4 \end{bmatrix}$. For the given vector $\mathbf{b}$, find the general form of all vectors $\mathbf{x}$ in R^3 for which $T_A(\mathbf{x}) = \mathbf{b}$ if such vectors exist.

a. $\mathbf{b} = (0, 0)$ **b.** $\mathbf{b} = (1, 3)$ **c.** $\mathbf{b} = (-1, 1)$

22. In each part, let $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 1 & 1 \\ 2 & 0 \end{bmatrix}$. For the given vector $\mathbf{b}$, find the general form of all vectors $\mathbf{x}$ in R^2 for which $T_A(\mathbf{x}) = \mathbf{b}$ if such vectors exist.

a. $\mathbf{b} = (0, 0, 0, 0)$ **b.** $\mathbf{b} = (1, 1, -1, -1)$
c. $\mathbf{b} = (2, 0, 0, 2)$

23. **a.** The equation $x + y + z = 1$ can be viewed as a linear system of one equation in three unknowns. Express a general solution of this equation as a particular solution plus a general solution of the associated homogeneous equation.

b. Give a geometric interpretation of the result in part (a).

24. **a.** The equation $x + y = 1$ can be viewed as a linear system of one equation in two unknowns. Express a general solution of this equation as a particular solution plus a general solution of the associated homogeneous system.

b. Give a geometric interpretation of the result in part (a).

25. Consider the linear systems

$$\begin{bmatrix} 3 & 2 & -1 \\ 6 & 4 & -2 \\ -3 & -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} 3 & 2 & -1 \\ 6 & 4 & -2 \\ -3 & -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix}$$

a. Find a general solution of the homogeneous system.

b. Confirm that $x_1 = 1, x_2 = 0, x_3 = 1$ is a solution of the nonhomogeneous system.

c. Use the results in parts (a) and (b) to find a general solution of the nonhomogeneous system.

d. Check your result in part (c) by solving the nonhomogeneous system directly.

26. Consider the linear systems

$$\begin{bmatrix} 1 & -2 & 3 \\ 2 & 1 & 4 \\ 1 & -7 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & -2 & 3 \\ 2 & 1 & 4 \\ 1 & -7 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 7 \\ -1 \end{bmatrix}$$

a. Find a general solution of the homogeneous system.

b. Confirm that $x_1 = 1, x_2 = 1, x_3 = 1$ is a solution of the nonhomogeneous system.

c. Use the results in parts (a) and (b) to find a general solution of the nonhomogeneous system.

d. Check your result in part (c) by solving the nonhomogeneous system directly.

In Exercises 27–28, find a general solution of the system, and use that solution to find a general solution of the associated homogeneous system and a particular solution of the given system.

27. $\begin{bmatrix} 3 & 4 & 1 & 2 \\ 6 & 8 & 2 & 5 \\ 9 & 12 & 3 & 10 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ 13 \end{bmatrix}$

28. $\begin{bmatrix} 9 & -3 & 5 & 6 \\ 6 & -2 & 3 & 1 \\ 3 & -1 & 3 & 14 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ -8 \end{bmatrix}$

29. a. Let

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

Show that relative to an *xyz*-coordinate system in 3-space the null space of A consists of all points on the *z*-axis and that the column space consists of all points in the *xy*-plane (see the accompanying figure).

b. Find a 3×3 matrix whose null space is the *x*-axis and whose column space is the *yz*-plane.

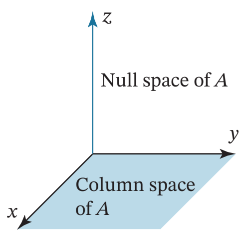


FIGURE Ex-29

30. Find a 3×3 matrix whose null space is

a. a point. **b.** a line. **c.** a plane.

31. a. Find all 2×2 matrices whose null space is the line $3x - 5y = 0$.

b. Describe the null spaces of the following matrices:

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

Working with Proofs

32. Prove Theorem 4.8.4.

33. Prove that the row vectors of an $n \times n$ invertible matrix A form a basis for R^n.

34. Suppose that A and B are $n \times n$ matrices and A is invertible. Invent and prove a theorem that describes how the row spaces of AB and B are related.

True-False Exercises

TF. In parts **(a)–(j)** determine whether the statement is true or false, and justify your answer.

a. The span of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is the column space of the matrix whose column vectors are $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

b. The column space of a matrix A is the set of solutions of $A\mathbf{x} = \mathbf{b}$.

c. If R is the reduced row echelon form of A, then those column vectors of R that contain the leading 1's form a basis for the column space of A.

d. The set of nonzero row vectors of a matrix A is a basis for the row space of A.

e. If A and B are $n \times n$ matrices that have the same row space, then A and B have the same column space.

f. If E is an $m \times m$ elementary matrix and A is an $m \times n$ matrix, then the null space of EA is the same as the null space of A.

g. If E is an $m \times m$ elementary matrix and A is an $m \times n$ matrix, then the row space of EA is the same as the row space of A.

h. If E is an $m \times m$ elementary matrix and A is an $m \times n$ matrix, then the column space of EA is the same as the column space of A.

i. The system $A\mathbf{x} = \mathbf{b}$ is inconsistent if and only if $\mathbf{b}$ is not in the column space of A.

j. There is an invertible matrix A and a singular matrix B such that the row spaces of A and B are the same.

Working with Technology

T1. Find a basis for the column space of

$$A = \begin{bmatrix} 2 & 6 & 0 & 8 & 4 & 12 & 4 \\ 3 & 9 & -2 & 8 & 6 & 18 & 6 \\ 3 & 9 & -7 & -2 & 6 & -3 & -1 \\ 2 & 6 & 5 & 18 & 4 & 33 & 11 \\ 1 & 3 & -2 & 0 & 2 & 6 & 2 \end{bmatrix}$$

that consists of column vectors of A.

T2. Find a basis for the row space of the matrix A in Exercise T1 that consists of row vectors of A.

4.9 Rank, Nullity, and the Fundamental Matrix Spaces

In the last section we investigated relationships between a system of linear equations and the row space, column space, and null space of its coefficient matrix. In this section we will be concerned with the dimensions of those spaces. The results we obtain will provide a deeper insight into the relationship between a linear system and its coefficient matrix.

Row and Column Spaces Have Equal Dimensions

In Examples 6 and 7 of Section 4.8 we found that the row and column spaces of the matrix

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

both have three basis vectors and hence are both three-dimensional. The fact that these spaces have the same dimension is not accidental, but rather a consequence of the following theorem.

Theorem 4.9.1

The row space and the column space of a matrix A have the same dimension.

Proof It follows from Theorems 4.8.4 and 4.8.6 (b) that elementary row operations do not change the dimension of the row space or of the column space of a matrix. Thus, if R is any row echelon form of A, it must be true that

$$\dim(\text{row space of } A) = \dim(\text{row space of } R)$$
$$\dim(\text{column space of } A) = \dim(\text{column space of } R)$$

so it suffices to show that the row and column spaces of R have the same dimension. But the dimension of the row space of R is the number of nonzero rows, and by Theorem 4.8.5 the dimension of the column space of R is the number of leading 1's. Since these two numbers are the same, the row and column space have the same dimension. ■

Rank and Nullity

The dimensions of the row space, column space, and null space of a matrix are such important numbers that there is some notation and terminology associated with them.

The proof of Theorem 4.9.1 shows that the rank of A can be interpreted as the number of leading 1's in any row echelon form of A.

Definition 1

The common dimension of the row space and column space of a matrix A is called the ***rank*** of A and is denoted by rank(A); the dimension of the null space of A is called the ***nullity*** of A and is denoted by nullity(A).

EXAMPLE 1 | Rank and Nullity of a 4×6 Matrix

Find the rank and nullity of the matrix

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

Solution The reduced row echelon form of A is

$$\begin{bmatrix} 1 & 0 & -4 & -28 & -37 & 13 \\ 0 & 1 & -2 & -12 & -16 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{1}$$

(verify). Since this matrix has two leading 1's, its row and column spaces are two-dimensional and rank$(A) = 2$. To find the nullity of A, we must find the dimension of the solution space of the linear system $A\mathbf{x} = \mathbf{0}$. This system can be solved by reducing its augmented matrix to reduced row echelon form. The resulting matrix will be identical to (1), except that it will have an additional last column of zeros, and hence the corresponding system of equations will be

$$\begin{aligned} x_1 - 4x_3 - 28x_4 - 37x_5 + 13x_6 &= 0 \\ x_2 - 2x_3 - 12x_4 - 16x_5 + 5x_6 &= 0 \end{aligned}$$

Solving these equations for the leading variables yields

$$\begin{aligned} x_1 &= 4x_3 + 28x_4 + 37x_5 - 13x_6 \\ x_2 &= 2x_3 + 12x_4 + 16x_5 - 5x_6 \end{aligned} \tag{2}$$

from which we obtain the general solution

$$\begin{aligned} x_1 &= 4r + 28s + 37t - 13u \\ x_2 &= 2r + 12s + 16t - 5u \\ x_3 &= r \\ x_4 &= s \\ x_5 &= t \\ x_6 &= u \end{aligned}$$

or in column vector form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = r\begin{bmatrix} 4 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} 28 \\ 12 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 37 \\ 16 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + u\begin{bmatrix} -13 \\ -5 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{3}$$

Because the four vectors on the right side of Formula (3) form a basis for the solution space it follows that nullity$(A) = 4$.

EXAMPLE 2 | Maximum Value for Rank

What is the maximum possible rank of an $m \times n$ matrix A that is not square?

Solution Since the row vectors of A lie in R^n and the column vectors in R^m, the row space of A is at most n-dimensional and the column space is at most m-dimensional. Since the rank of A is the common dimension of its row and column space, it follows that the rank is at most the smaller of m and n. We denote this by writing

$$\text{rank}(A) \le \min(m, n)$$

in which $\min(m, n)$ is the minimum of m and n.

The following theorem establishes a fundamental relationship between the rank and nullity of a matrix.

Theorem 4.9.2

Dimension Theorem for Matrices

If A is a matrix with n columns, then

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n \tag{4}$$

Proof Since A has n columns, the homogeneous linear system $A\mathbf{x} = \mathbf{0}$ has n unknowns (variables). These fall into two distinct categories: the leading variables and the free variables. Thus,

$$\begin{bmatrix}\text{number of leading}\\ \text{variables}\end{bmatrix} + \begin{bmatrix}\text{number of free}\\ \text{variables}\end{bmatrix} = n$$

But the number of leading variables is the same as the number of leading 1's in any row echelon form of A, which is the same as the dimension of the row space of A, which is the same as the rank of A. Also, the number of free variables in the general solution of $A\mathbf{x} = \mathbf{0}$ is the same as the number of parameters in that solution, which is the same as the dimension of the solution space of $A\mathbf{x} = \mathbf{0}$, which is the same as the nullity of A. This yields Formula (4). ■

EXAMPLE 3 | The Sum of Rank and Nullity

The matrix

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

has 6 columns, so

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = 6$$

This is consistent with Example 1, where we showed that

$$\operatorname{rank}(A) = 2 \quad \text{and} \quad \operatorname{nullity}(A) = 4$$

The following theorem, which summarizes results already obtained, interprets rank and nullity in the context of a homogeneous linear system.

Theorem 4.9.3

If A is an $m \times n$ matrix, then

(*a*) $\operatorname{rank}(A)$ = the number of leading variables in the general solution of $A\mathbf{x} = \mathbf{0}$.

(*b*) $\operatorname{nullity}(A)$ = the number of parameters in the general solution of $A\mathbf{x} = \mathbf{0}$.

EXAMPLE 4 | Rank, Nullity, and Linear Systems

(*a*) Find the number of parameters in the general solution of $A\mathbf{x} = \mathbf{0}$ if A is a 5×7 matrix of rank 3.

(*b*) Find the rank of a 5×7 matrix A for which $A\mathbf{x} = \mathbf{0}$ has a two-dimensional solution space.

Solution (*a*) From (4),

$$\text{nullity}(A) = n - \text{rank}(A) = 7 - 3 = 4$$

Thus, there are four parameters.

Solution (*b*) The matrix A has nullity 2, so

$$\text{rank}(A) = n - \text{nullity}(A) = 7 - 2 = 5$$

Recall from Section 4.8 that if $A\mathbf{x} = \mathbf{b}$ is a consistent linear system, then its general solution can be expressed as the sum of a particular solution of this system and the general solution of $A\mathbf{x} = \mathbf{0}$. We leave it as an exercise for you to use this fact and Theorem 4.9.3 to prove the following result.

Theorem 4.9.4

If $A\mathbf{x} = \mathbf{b}$ is a consistent linear system of m equations in n unknowns, and if A has rank r, then the general solution of the system contains $n - r$ parameters.

The Fundamental Spaces of a Matrix

There are six important vector spaces associated with an $m \times n$ matrix A and its transpose A^T:

row space of A	row space of A^T
column space of A	column space of A^T
null space of A	null space of A^T

However, transposing a matrix converts row vectors into column vectors and conversely, so except for a difference in notation, the row space of A^T is the same as the column space of A, and the column space of A^T is the same as the row space of A. Thus, of the six spaces listed above, only the following four are distinct:

row space of A	column space of A
null space of A	null space of A^T

These are called the ***fundamental spaces*** of the matrix A. The row space and null space of A are subspaces of R^n, whereas the column space of A and the null space of A^T are subspaces of R^m. The null space of A^T is also called the ***left null space of A*** because transposing both sides of the equation $A^T\mathbf{x} = \mathbf{0}$ produces the equation $\mathbf{x}^T A = \mathbf{0}^T$ in which the unknown is on the left. The dimension of the left null space of A is called the ***left nullity of A***. We will now consider how the four fundamental spaces are related.

Let us focus for a moment on the matrix A^T. Since the row space and column space of a matrix have the same dimension, and since transposing a matrix converts its columns to rows and its rows to columns, the following result should not be surprising.

Theorem 4.9.5

If A is any matrix, then $\operatorname{rank}(A) = \operatorname{rank}(A^T)$.

Proof

$$\operatorname{rank}(A) = \dim(\text{row space of } A) = \dim(\text{column space of } A^T) = \operatorname{rank}(A^T). \quad \blacksquare$$

This result has some important implications. For example, if A is an $m \times n$ matrix, then applying Formula (4) to the matrix A^T and using the fact that this matrix has m columns yields

$$\operatorname{rank}(A^T) + \operatorname{nullity}(A^T) = m$$

which, by virtue of Theorem 4.9.5, can be rewritten as

$$\boxed{\operatorname{rank}(A) + \operatorname{nullity}(A^T) = m} \tag{5}$$

This alternative form of Formula (4) makes it possible to express the dimensions of all four fundamental spaces in terms of the size and rank of A. Specifically, if $\operatorname{rank}(A) = r$, then

$$\boxed{\begin{array}{ll} \dim[\operatorname{row}(A)] = r & \dim[\operatorname{col}(A)] = r \\ \dim[\operatorname{null}(A)] = n - r & \dim[\operatorname{null}(A^T)] = m - r \end{array}} \tag{6}$$

Bases for the Fundamental Spaces

An efficient way to obtain bases for the four fundamental spaces of an $m \times n$ matrix A is to adjoin the $m \times m$ identity matrix to A to obtain an augmented matrix $[A \mid I]$ and apply elementary row operations to this matrix to put A in reduced row echelon form R, thereby putting the augmented matrix in the form $[R \mid E]$. In the case where A is invertible the matrix E will be A^{-1}, but in general it will not. The rank r of A can then be obtained by counting the number of pivots (leading l's) in R, and the nullity of A^T can be obtained from the relationship

$$\operatorname{nullity}(A^T) = m - r \tag{7}$$

that follows from Formula (5). Bases for three of the fundamental spaces can be obtained directly from $[R \mid E]$ as follows:

- A basis for $\operatorname{row}(A)$ will be the r rows of R that contain the leading 1's (the pivot rows).
- A basis for $\operatorname{col}(A)$ will be the r columns of A that contain the leading 1's of R (the pivot columns).
- A basis for $\operatorname{null}(A^T)$ will be the bottom $m - r$ rows of E (see the proof at the end of this section)

EXAMPLE 5 | Bases for the Fundamental Spaces

In Example 1 we found a basis for the null space of the 4×6 matrix

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

so in this example we will focus on finding bases for the remaining three fundamental spaces starting with the matrix

$$\underbrace{\left[\begin{array}{rrrrrr|rrrr} -1 & 2 & 0 & 4 & 5 & -3 & 1 & 0 & 0 & 0 \\ 3 & -7 & 2 & 0 & 1 & 4 & 0 & 1 & 0 & 0 \\ 2 & -5 & 2 & 4 & 6 & 1 & 0 & 0 & 1 & 0 \\ 4 & -9 & 2 & -4 & -4 & 7 & 0 & 0 & 0 & 1 \end{array}\right]}_{A \qquad\qquad\qquad I}$$

in which a 4×4 identity matrix has been adjoined to A. Using Gaussian elimination to reduce the left side to reduced row echelon form R yields (verify)

$$\underbrace{\left[\begin{array}{rrrrrr|rrrr} 1 & 0 & -4 & -28 & -37 & 13 & 0 & 0 & -\frac{9}{2} & \frac{5}{2} \\ 0 & 1 & -2 & -12 & -16 & 5 & 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\frac{1}{2} & -\frac{1}{2} \end{array}\right]}_{R \qquad\qquad\qquad E}$$

From R we see that A has rank $r = 2$ (two nonzero rows), has nullity $n - r = 6 - 2 = 4$, and from (7) has left nullity $m - r = 2$. The two pivot rows of R (rows 1 and 2) form a basis for the row space of A, the two pivot columns of A (columns 1 and 2) form a basis for the column space of A, and the bottom two rows of E form a basis for the left null space of A. Expressing these bases in column form we have:

$$\text{row space basis: } \left\{ \begin{bmatrix} -1 \\ 0 \\ -4 \\ -28 \\ -37 \\ 13 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -2 \\ -12 \\ -16 \\ 5 \end{bmatrix} \right\}, \quad \text{column space basis: } \left\{ \begin{bmatrix} -1 \\ 3 \\ 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ -7 \\ -5 \\ -9 \end{bmatrix} \right\}$$

$$\text{left null space basis: } \left\{ \begin{bmatrix} 1 \\ 0 \\ -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} \right\}$$

A Geometric Link Between the Fundamental Spaces

The four formulas in (6) provide an *algebraic* relationship between the size of a matrix and the dimensions of its fundamental spaces. Our next objective is to find a *geometric* relationship between the fundamental spaces themselves. For this purpose recall from Theorem 3.4.3 that if A is an $m \times n$ matrix, then the null space of A consists of those vectors that are orthogonal to each of the row vectors of A. To develop that idea in more detail, we make the following definition.

Definition 2

If W is a subspace of R^n, then the set of all vectors in R^n that are orthogonal to every vector in W is called the ***orthogonal complement*** of W and is denoted by the symbol $W^{\perp}$.

The following theorem lists three basic properties of orthogonal complements. We will omit the formal proof because a more general version of this theorem will be proved later in the text.

Theorem 4.9.6

If W is a subspace of R^n, then:

(a) $W^\perp$ is a subspace of R^n.
(b) The only vector common to W and $W^\perp$ is $\mathbf{0}$.
(c) The orthogonal complement of $W^\perp$ is W.

Part (b) of Theorem 4.9.6 can be expressed as

$$W \cap W^\perp = \{\mathbf{0}\}$$

and part (c) as

$$(W^\perp)^\perp = W$$

EXAMPLE 6 | Orthogonal Complements

In R^2 the orthogonal complement of a line W through the origin is the line through the origin that is perpendicular to W (**Figure 4.9.1*a***); and in R^3 the orthogonal complement of a plane W through the origin is the line through the origin that is perpendicular to that plane (**Figure 4.9.1*b***).

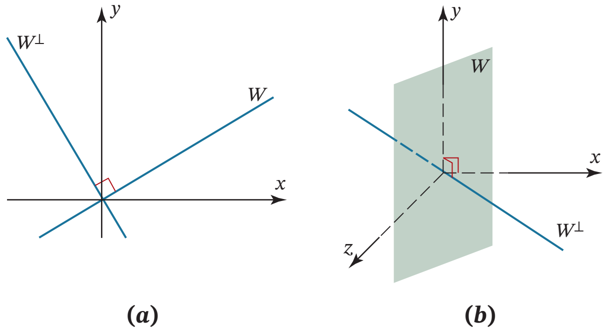


FIGURE 4.9.1

The next theorem will provide a geometric link between the fundamental spaces of a matrix. In the exercises we will ask you to prove that if a vector in R^n is orthogonal to each vector in a *basis* for a subspace of R^n, then it is orthogonal to *every* vector in that subspace. Thus, part (a) of the following theorem is essentially a restatement of Theorem 3.4.3 in the language of orthogonal complements; it is illustrated in Example 6 of Section 3.4. The proof of part (b), which is left as an exercise, follows from part (a). The essential idea of the theorem is illustrated in **Figure 4.9.2**.

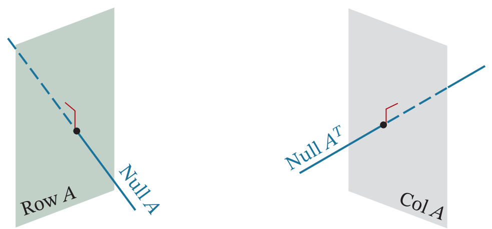


FIGURE 4.9.2

Theorem 4.9.7

If A is an $m \times n$ matrix, then:

(a) The null space of A and the row space of A are orthogonal complements in R^n.
(b) The null space of A^T and the column space of A are orthogonal complements in R^m.

Explain why $\{\mathbf{0}\}$ and R^n are orthogonal complements.

The results in Theorem 4.9.7 are often illustrated as in **Figure 4.9.3**, which conveys the orthogonality properties in the theorem as well as the dimensions of the fundamental spaces.

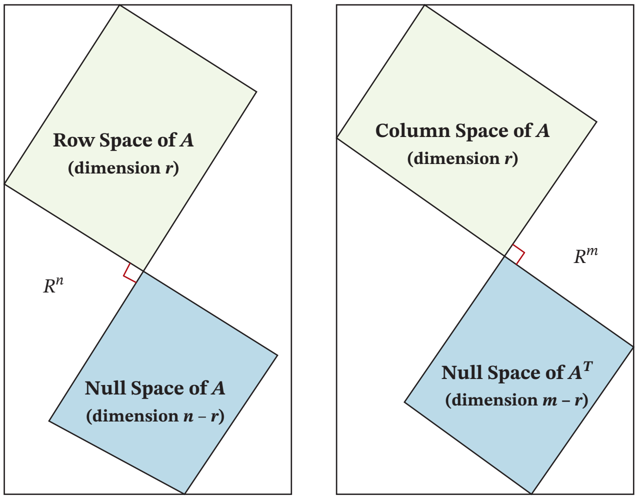


FIGURE 4.9.3

More on the Equivalence Theorem

In Theorem 2.3.8 we listed seven results that are equivalent to the invertibility of a square matrix A. We are now in a position to add ten more statements to that list to produce a single theorem that summarizes and links together all of the topics that we have covered thus far. We will prove some of the equivalences and leave others as exercises.

Theorem 4.9.8

Equivalent Statements

If A is an $n \times n$ matrix in which there are no duplicate rows and no duplicate columns, then the following statements are equivalent.

(*a*) A is invertible.

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

(*c*) The reduced row echelon form of A is I_n.

(*d*) A is expressible as a product of elementary matrices.

(*e*) $A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.

(*f*) $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.

(*g*) $\det(A) \neq 0$.

(*h*) The column vectors of A are linearly independent.

(*i*) The row vectors of A are linearly independent.

(*j*) The column vectors of A span R^n.

(*k*) The row vectors of A span R^n.

(*l*) The column vectors of A form a basis for R^n.

(*m*) The row vectors of A form a basis for R^n.

(*n*) A has rank n.

(*o*) A has nullity 0.

(*p*) The orthogonal complement of the null space of A is R^n.

(*q*) The orthogonal complement of the row space of A is $\{\mathbf{0}\}$.

Proof The following proofs show that (b) implies (h) through (q). In the exercises we will ask you to complete the proof by showing that (q) implies (b).

$(b) \Rightarrow (h)$ By Formula (10) of Section 1.3, $A\mathbf{x}$ is a linear combination of the column vectors of A. Since $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, the column vectors of A must be linearly independent.

$(h) \Rightarrow (j)$, $(h) \Rightarrow (l)$, $(h) \Rightarrow (n)$ Since we now know that the n column vectors of A are linearly independent vectors in the n-dimensional vector space R^n, they must span R^n by Theorem 4.6.4 and hence form a basis for R^n. This also means that $\text{rank}(A) = n$.

$(h) \Rightarrow (i)$, $(h) \Rightarrow (k)$, $(h) \Rightarrow (m)$ Since we have shown that the column vectors form a basis for R^n, and since the row space and column space of A have the same dimension by Theorem 4.9.1, the n row vectors of A must also form a basis for R^n.

$(n) \Rightarrow (o)$ Since $\text{rank}(A) = n$, it follows from Theorem 4.9.2 that $\text{nullity}(A) = 0$.

$(o) \Rightarrow (p)$ $\text{nullity}(A) = 0$ means that the null space of A is $\{\mathbf{0}\}$, and since every vector in R^n is orthogonal to $\mathbf{0}$, it follows that the orthogonal complement of the null space of A is R^n.

$(p) \Rightarrow (q)$ It follows from Theorem 4.9.7 that orthogonal complement of the row space of A is the null space of A, which is $\{\mathbf{0}\}$. ■

Applications of Rank

The advent of the Internet has stimulated research on finding efficient methods for transmitting large amounts of digital data over communications lines with limited bandwidths. Digital data are commonly stored in matrix form, and many techniques for improving transmission speed use the rank of a matrix in some way. Rank plays a role because it measures the "redundancy" in a matrix in the sense that if A is an $m \times n$ matrix of rank k, then $n - k$ of the column vectors and $m - k$ of the row vectors can be expressed in terms of k linearly independent column or row vectors. The essential idea in many data compression schemes is to approximate the original data set by a data set with smaller rank that conveys nearly the same information, then eliminate redundant vectors in the approximating set to speed up the transmission time.

OPTIONAL: Overdetermined and Underdetermined Systems

In many applications the equations in a linear system correspond to physical constraints or conditions that must be satisfied. In general, the most desirable systems are those that have the same number of constraints as unknowns since such systems often have a unique solution. Unfortunately, it is not always possible to match the number of constraints and unknowns, so researchers are often faced with linear systems that have more constraints than unknowns, called ***overdetermined systems***, or with fewer constraints than unknowns, called ***underdetermined systems***. The following theorem will help us to analyze both overdetermined and underdetermined systems.

Theorem 4.9.9

Let A be an $m \times n$ matrix.

(a) (***Overdetermined Case***). If $m > n$, then the linear system $A\mathbf{x} = \mathbf{b}$ is inconsistent for at least one vector $\mathbf{b}$ in R^n.

(b) (***Underdetermined Case***). If $m < n$, then for each vector $\mathbf{b}$ in R^m the linear system $A\mathbf{x} = \mathbf{b}$ is either inconsistent or has infinitely many solutions.

In engineering and physics, the occurrence of an overdetermined or underdetermined linear system often signals that one or more variables were omitted in formulating the problem or that extraneous variables were included. This often leads to some kind of complication.

Proof (a) Assume that $m > n$, in which case the column vectors of A cannot span R^m (fewer vectors than the dimension of R^m). Thus, there is at least one vector $\mathbf{b}$ in R^m that is not in the column space of A, and for any such $\mathbf{b}$ the system $A\mathbf{x} = \mathbf{b}$ is inconsistent by Theorem 4.8.1.

Proof (b) Assume that $m < n$. For each vector $\mathbf{b}$ in R^n there are two possibilities: either the system $A\mathbf{x} = \mathbf{b}$ is consistent or it is inconsistent. If it is inconsistent, then the proof is complete. If it is consistent, then Theorem 4.9.4 implies that the general solution has $n - r$ parameters, where $r = \text{rank}(A)$. But we know from Example 2 that rank(A) is at most the smaller of m and n (which is m), so

$$n - r \geq n - m > 0$$

This means that the general solution has at least one parameter and hence there are infinitely many solutions. ■

EXAMPLE 7 | Overdetermined and Underdetermined Systems

(a) What can you say about the solutions of an overdetermined system $A\mathbf{x} = \mathbf{b}$ of 7 equations in 5 unknowns in which A has rank $r = 4$?

(b) What can you say about the solutions of an underdetermined system $A\mathbf{x} = \mathbf{b}$ of 5 equations in 7 unknowns in which A has rank $r = 4$?

Solution (a) The system is consistent for some vector $\mathbf{b}$ in R^7, and for any such $\mathbf{b}$ the number of parameters in the general solution is $n - r = 5 - 4 = 1$.

Solution (b) The system may be consistent or inconsistent, but if it is consistent for the vector $\mathbf{b}$ in R^5, then the general solution has $n - r = 7 - 4 = 3$ parameters.

EXAMPLE 8 | An Overdetermined System

The linear system

$$\begin{aligned} x_1 - 2x_2 &= b_1 \\ x_1 - x_2 &= b_2 \\ x_1 + x_2 &= b_3 \\ x_1 + 2x_2 &= b_4 \\ x_1 + 3x_2 &= b_5 \end{aligned}$$

is overdetermined, so it cannot be consistent for all possible values of b_1, b_2, b_3, b_4, and b_5. Conditions under which the system is consistent can be obtained by solving the linear system by Gauss–Jordan elimination. We leave it for you to show that the augmented matrix is row equivalent to

$$\begin{bmatrix} 1 & 0 & 2b_2 - b_1 \\ 0 & 1 & b_2 - b_1 \\ 0 & 0 & b_3 - 3b_2 + 2b_1 \\ 0 & 0 & b_4 - 4b_2 + 3b_1 \\ 0 & 0 & b_5 - 5b_2 + 4b_1 \end{bmatrix} \tag{8}$$

Thus, the system is consistent if and only if b_1, b_2, b_3, b_4, and b_5 satisfy the conditions

$$\begin{aligned} 2b_1 - 3b_2 + b_3 \qquad\qquad &= 0 \\ 3b_1 - 4b_2 \qquad + b_4 \qquad &= 0 \\ 4b_1 - 5b_2 \qquad\qquad + b_5 &= 0 \end{aligned}$$

Solving this homogeneous linear system yields

$$b_1 = 5r - 4s, \quad b_2 = 4r - 3s, \quad b_3 = 2r - s, \quad b_4 = r, \quad b_5 = s$$

where r and s are arbitrary.

Remark The coefficient matrix for the given linear system in the last example has $n = 2$ columns, and it has rank $r = 2$ because there are two nonzero rows in its reduced row echelon form. This implies that when the system is consistent its general solution will contain $n - r = 0$ parameters; that is, the solution will be unique. With a moment's thought, you should be able to see that this is so from (8).

OPTIONAL: Left Null Space Proof

Suppose that A is an $m \times n$ matrix of rank r and its reduced row echelon form is R. We will conclude this section by proving that if the augmented matrix $[A \mid I]$ is reduced to $[R \mid E]$ by Gauss-Jordan elimination, then the bottom $m - r$ rows of E form a basis for the left null space of A.

Proof The left null space of A is the solution space of the system $A^T\mathbf{x} = \mathbf{0}$, which, on transposing both sides, we can rewrite as

$$\mathbf{x}^T A = \mathbf{0}^T \tag{9}$$

Let $[R \mid E]$ denote the augmented matrix that results from $[A \mid I]$, when elementary row operations are applied to put the left side in reduced row echelon form R. The matrices A, R, and E are related by the equation

$$EA = R$$

where E is a product of elementary matrices. Since A has rank r and size $m \times n$, the matrix R has r nonzero rows and $m - r$ zero rows. By Formula (9) of Section 1.3 the ith row vector of R is the product

$$[i\text{th row vector of } E]\ A = i\text{th row vector of } R$$

But the last $m - r$ row vectors of R are zero, so the last $m - r$ row vectors of E are solutions of (9) and hence lie in the left null space of A. We leave it as an exercise to use Theorem 4.9.8 to show that these vectors form a basis for the left null space of A. ■

Exercise Set 4.9

In Exercises 1–2, find the rank and nullity of the matrix A by reducing it to row echelon form.

1. a. $A = \begin{bmatrix} 1 & 2 & -1 & 1 \\ 2 & 4 & -2 & 2 \\ 3 & 6 & -3 & 3 \\ 4 & 8 & -4 & 4 \end{bmatrix}$

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

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

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

In Exercises 3–6, the matrix R is the reduced row echelon form of the matrix A.

a. *By inspection of the matrix R, find the rank and nullity of A.*

b. *Confirm that the rank and nullity satisfy Formula (4).*

c. *Find the number of leading variables and the number of parameters in the general solution of $A\mathbf{x} = \mathbf{0}$ without solving the system.*

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

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

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

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

7. In each part, find the largest possible value for the rank of A and the smallest possible value for the nullity of A.

a. A is 4×4 **b.** A is 3×5 **c.** A is 5×3

8. If A is an $m \times n$ matrix, what is the largest possible value for its rank and the smallest possible value for its nullity?

9. In each part, use the information in the table to:

i. find the dimensions of the row space of A, column space of A, null space of A, and null space of A^T;

ii. determine whether the linear system $A\mathbf{x} = \mathbf{b}$ is consistent;

iii. find the number of parameters in the general solution of each system in (ii) that is consistent.

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Size of A	3×3	3×3	3×3	5×9	5×9	4×4	6×2
Rank(A)	3	2	1	2	2	0	2
Rank[$A \mid \mathbf{b}$]	3	3	1	2	3	0	2

10. Verify that $\operatorname{rank}(A) = \operatorname{rank}(A^T)$.

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

In Exercises 11–14 find the dimensions and bases for the four fundamental spaces of the matrix.

11. $A = \begin{bmatrix} 1 & 4 \\ 0 & 3 \\ -9 & 0 \end{bmatrix}$

12. $A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \end{bmatrix}$

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

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

In Exercises 15–18 confirm the orthogonality statements in the two parts of Theorem 4.9.7 for the given matrix.

15. The matrix in Exercise 11.

16. The matrix in Exercise 12.

17. The matrix in Exercise 13.

18. The matrix in Exercise 14.

In Exercises 19–20 use the method of Example 5 to find bases for the four fundamental spaces of the matrix.

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

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

21. a. Find an equation relating $\operatorname{nullity}(A)$ and $\operatorname{nullity}(A^T)$ for the matrix in Exercise 10.

b. Find an equation relating $\operatorname{nullity}(A)$ and $\operatorname{nullity}(A^T)$ for a general $m \times n$ matrix.

22. Let $T: R^2 \to R^3$ be the linear transformation defined by the formula

$$T(x_1, x_2) = (x_1 + 3x_2, x_1 - x_2, x_1)$$

a. Find the rank of the standard matrix for T.

b. Find the nullity of the standard matrix for T.

23. Let $T: R^5 \to R^3$ be the linear transformation defined by the formula

$$T(x_1, x_2, x_3, x_4, x_5) = (x_1 + x_2, x_2 + x_3 + x_4, x_4 + x_5)$$

a. Find the rank of the standard matrix for T.

b. Find the nullity of the standard matrix for T.

24. Discuss how the rank of A varies with t.

a. $A = \begin{bmatrix} 1 & 1 & t \\ 1 & t & 1 \\ t & 1 & 1 \end{bmatrix}$

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

25. Are there values of r and s for which

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

has rank 1? Has rank 2? If so, find those values.

26. a. Give an example of a 3×3 matrix whose column space is a plane through the origin in 3-space.

b. What kind of geometric object is the null space of your matrix?

c. What kind of geometric object is the row space of your matrix?

27. Suppose that A is a 3×3 matrix whose null space is a line through the origin in 3-space. Can the row or column space of A also be a line through the origin? Explain.

28. a. If A is a 3×5 matrix, then the rank of A is at most ________. Why?

b. If A is a 3×5 matrix, then the nullity of A is at most ________. Why?

c. If A is a 3×5 matrix, then the rank of A^T is at most ________. Why?

d. If A is a 3×5 matrix, then the nullity of A^T is at most ________. Why?

29. a. If A is a 3×5 matrix, then the number of leading 1's in the reduced row echelon form of A is at most ________. Why?

b. If A is a 3×5 matrix, then the number of parameters in the general solution of $A\mathbf{x} = \mathbf{0}$ is at most ________. Why?

c. If A is a 5×3 matrix, then the number of leading 1's in the reduced row echelon form of A is at most ________. Why?

d. If A is a 5×3 matrix, then the number of parameters in the general solution of $A\mathbf{x} = \mathbf{0}$ is at most ________. Why?

30. Let A be a 7×6 matrix such that $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Find the rank and nullity of A.

31. Let A be a 5×7 matrix with rank 4.

a. What is the dimension of the solution space of $A\mathbf{x} = \mathbf{0}$?

b. Is $A\mathbf{x} = \mathbf{b}$ consistent for all vectors $\mathbf{b}$ in R^5? Explain.

32. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

Show that A has rank 2 if and only if one or more of the following determinants is nonzero.

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}, \quad \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

33. Use the result in Exercise 22 to show that the set of points (x, y, z) in R^3 for which the matrix

$$\begin{bmatrix} x & y & z \\ 1 & x & y \end{bmatrix}$$

has rank 1 is the curve with parametric equations $x = t, y = t^2, z = t^3$.

34. Find matrices A and B for which $\operatorname{rank}(A) = \operatorname{rank}(B)$, but $\operatorname{rank}(A^2) \neq \operatorname{rank}(B^2)$.

35. In Example 6 of Section 4.7 we showed that the row space and the null space of the matrix

$$A = \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 \\ 0 & 0 & 5 & 10 & 0 & 15 \\ 2 & 6 & 0 & 8 & 4 & 18 \end{bmatrix}$$

are orthogonal complements in R^6, as guaranteed by part (a) of Theorem 4.9.7. Show that null space of A^T and the column space of A are orthogonal complements in R^4, as guaranteed by part (b) of Theorem 4.9.7. [*Suggestion:* Show that each column vector of A is orthogonal to each vector in a basis for the null space of A^T.]

36. Confirm the results stated in Theorem 4.9.7 for the matrix.

$$A = \begin{bmatrix} -2 & -5 & 8 & 0 & -17 \\ 1 & 3 & -5 & 1 & 5 \\ 3 & 11 & -19 & 7 & 1 \\ 1 & 7 & -13 & 5 & -3 \end{bmatrix}$$

37. In each part, state whether the system is overdetermined or underdetermined. If overdetermined, find all values of the b's for which it is inconsistent, and if underdetermined, find all values of the b's for which it is inconsistent and all values for which it has infinitely many solutions.

a. $\begin{bmatrix} 1 & -1 \\ -3 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$

b. $\begin{bmatrix} 1 & -3 & 4 \\ -2 & -6 & 8 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$

c. $\begin{bmatrix} 1 & -3 & 0 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$

38. What conditions must be satisfied by b_1, b_2, b_3, b_4, and b_5 for the overdetermined linear system

$$\begin{aligned} x_1 - 3x_2 &= b_1 \\ x_1 - 2x_2 &= b_2 \\ x_1 + x_2 &= b_3 \\ x_1 - 4x_2 &= b_4 \\ x_1 + 5x_2 &= b_5 \end{aligned}$$

to be consistent?

Working with Proofs

39. Prove: If $k \neq 0$, then A and kA have the same rank.

40. Prove: If a matrix A is not square, then either the row vectors or the column vectors of A are linearly dependent.

41. Use Theorem 4.9.3 to prove Theorem 4.9.4.

42. Prove Theorem 4.9.7(b).

43. Prove: If a vector $\mathbf{v}$ in R^n is orthogonal to each vector in a basis for a subspace W of R^n, then $\mathbf{v}$ is orthogonal to every vector in W.

44. Prove: (q) implies (b) in Theorem 4.9.8.

True-False Exercises

TF. In parts **(a)–(j)** determine whether the statement is true or false, and justify your answer.

a. Either the row vectors or the column vectors of a square matrix are linearly independent.

b. A matrix with linearly independent row vectors and linearly independent column vectors is square.

c. The nullity of a nonzero $m \times n$ matrix is at most m.

d. Adding one additional column to a matrix increases its rank by one.

e. The nullity of a square matrix with linearly dependent rows is at least one.

f. If A is square and $A\mathbf{x} = \mathbf{b}$ is inconsistent for some vector $\mathbf{b}$, then the nullity of A is zero.

g. If a matrix A has more rows than columns, then the dimension of the row space is greater than the dimension of the column space.

h. If $\text{rank}(A^T) = \text{rank}(A)$, then A is square.

i. There is no 3×3 matrix whose row space and null space are both lines in 3-space.

j. If V is a subspace of R^n and W is a subspace of V, then $W^{\perp}$ is a subspace of $V^{\perp}$.

Working with Technology

T1. It can be proved that a nonzero matrix A has rank k if and only if some $k \times k$ submatrix has a nonzero determinant and all square submatrices of larger size have determinant zero. Use this fact to find the rank of

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

Check your result by computing the rank of A in a different way.

T2. ***Sylvester's inequality*** states that if A and B are $n \times n$ matrices with rank r_A and r_B, respectively, then the rank r_{AB} of AB satisfies the inequality

$$r_A + r_B - n \leq r_{AB} \leq \min(r_A, r_B)$$

where $\min(r_A, r_B)$ denotes the smaller of r_A and r_B or their common value if the two ranks are the same. Use your technology utility to confirm this result for some matrices of your choice.

Chapter 4 Supplementary Exercises

1. Let V be the set of all ordered triples of real numbers, and consider the following addition and scalar multiplication operations on $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$:

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, u_3 + v_3), \quad k\mathbf{u} = (ku_1, 0, 0)$$

a. Compute $\mathbf{u} + \mathbf{v}$ and $k\mathbf{u}$ for $\mathbf{u} = (3, -2, 4)$, $\mathbf{v} = (1, 5, -2)$, and $k = -1$.

b. In words, explain why V is closed under addition and scalar multiplication.

c. Since the addition operation on V is the standard addition operation on R^3, certain vector space axioms hold for V because they are known to hold for R^3. Which axioms in Definition 1 of Section 4.1 are they?

d. Show that Axioms 7, 8, and 9 hold.

e. Show that Axiom 10 fails for the given operations.

2. In each part, the solution space of the system is a subspace of R^3 and so must be a line through the origin, a plane through the origin, all of R^3, or the origin only. For each system, determine which is the case. If the subspace is a plane, find an equation for it, and if it is a line, find parametric equations.

a. $0x + 0y + 0z = 0$

b.
$$\begin{aligned} 2x - 3y + z &= 0 \\ 6x - 9y + 3z &= 0 \\ -4x + 6y - 2z &= 0 \end{aligned}$$

c.
$$\begin{aligned} x - 2y + 7z &= 0 \\ -4x + 8y + 5z &= 0 \\ 2x - 4y + 3z &= 0 \end{aligned}$$

d.
$$\begin{aligned} x + 4y + 8z &= 0 \\ 2x + 5y + 6z &= 0 \\ 3x + y - 4z &= 0 \end{aligned}$$

3. For what values of s is the solution space of

$$\begin{aligned} x_1 + x_2 + sx_3 &= 0 \\ x_1 + sx_2 + x_3 &= 0 \\ sx_1 + x_2 + x_3 &= 0 \end{aligned}$$

the origin only, a line through the origin, a plane through the origin, or all of R^3?

4. **a.** Express $(4a, a - b, a + 2b)$ as a linear combination of $(4, 1, 1)$ and $(0, -1, 2)$.

b. Express $(3a + b + 3c, -a + 4b - c, 2a + b + 2c)$ as a linear combination of $(3, -1, 2)$ and $(1, 4, 1)$.

c. Express $(2a - b + 4c, 3a - c, 4b + c)$ as a linear combination of three nonzero vectors.

5. Let W be the space spanned by $\mathbf{f} = \sin x$ and $\mathbf{g} = \cos x$.

a. Show that for any value of θ, $\mathbf{f}_1 = \sin(x + \theta)$ and $\mathbf{g}_1 = \cos(x + \theta)$ are vectors in W.

b. Show that $\mathbf{f}_1$ and $\mathbf{g}_1$ form a basis for W.

6. **a.** Express $\mathbf{v} = (1, 1)$ as a linear combination of $\mathbf{v}_1 = (1, -1)$, $\mathbf{v}_2 = (3, 0)$, and $\mathbf{v}_3 = (2, 1)$ in two different ways.

b. Explain why this does not violate Theorem 4.5.1.

7. Let A be an $n \times n$ matrix, and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be linearly independent vectors in R^n expressed as $n \times 1$ matrices. What must be true about A for $A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_n$ to be linearly independent?

8. Must a basis for P_n contain a polynomial of degree k for each $k = 0, 1, 2, \ldots, n$? Justify your answer.

9. For the purpose of this exercise, let us define a "checkerboard matrix" to be a square matrix $A = [a_{ij}]$ such that

$$a_{ij} = \begin{cases} 1 & \text{if } i + j \text{ is even} \\ 0 & \text{if } i + j \text{ is odd} \end{cases}$$

Find the rank and nullity of the following checkerboard matrices.

a. The 3×3 checkerboard matrix.

b. The 4×4 checkerboard matrix.

c. The $n \times n$ checkerboard matrix.

10. For the purpose of this exercise, let us define an "X-matrix" to be a square matrix with an odd number of rows and columns that has 0's everywhere except on the two diagonals where it has 1's. Find the rank and nullity of the following X-matrices.

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

c. the X-matrix of size $(2n + 1) \times (2n + 1)$

11. In each part, show that the stated set of polynomials is a subspace of P_n and find a basis for it.

a. All polynomials in P_n such that $p(-x) = p(x)$.

b. All polynomials in P_n such that $p(0) = p(1)$.

12. ***(Calculus required)*** Show that the set of all polynomials in P_n that have a horizontal tangent at $x = 0$ is a subspace of P_n. Find a basis for this subspace.

13. a. Find a basis for the vector space of all 3×3 symmetric matrices.

b. Find a basis for the vector space of all 3×3 skew-symmetric matrices.

14. Various advanced texts in linear algebra prove the following determinant criterion for rank: *The rank of a matrix A is r if and only if A has some $r \times r$ submatrix with a nonzero determinant, and all square submatrices of larger size have determinant zero.* [*Note:* A submatrix of A is any matrix obtained by deleting rows or columns of A. The matrix A itself is also considered to be a submatrix of A.] In each part, use this criterion to find the rank of the matrix.

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

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

15. Use the result in Exercise 14 to find the possible ranks for matrices of the form

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & a_{16} \\ 0 & 0 & 0 & 0 & 0 & a_{26} \\ 0 & 0 & 0 & 0 & 0 & a_{36} \\ 0 & 0 & 0 & 0 & 0 & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \end{bmatrix}$$

16. Prove: If S is a basis for a vector space V, then for any vectors $\mathbf{u}$ and $\mathbf{v}$ in V and any scalar k, the following relationships hold.

a. $(\mathbf{u} + \mathbf{v})_S = (\mathbf{u})_S + (\mathbf{v})_S$ **b.** $(k\mathbf{u})_S = k(\mathbf{u})_S$

17. Let D_k, R_θ, and S_k be a dilation of R^2 with factor k, a counterclockwise rotation about the origin of R^2 through an angle θ, and a shear of R^2 by a factor k, respectively.

a. Do D_k and R_θ commute?

b. Do R_θ and S_k commute?

c. Do D_k and S_k commute?

18. A vector space V is said to be the ***direct sum*** of its subspaces U and W, written $V = U \oplus W$, if every vector in V can be expressed in exactly one way as $\mathbf{v} = \mathbf{u} + \mathbf{w}$, where $\mathbf{u}$ is a vector in U and $\mathbf{w}$ is a vector in W.

a. Prove that $V = U \oplus W$ if and only if every vector in V is the sum of some vector in U and some vector in W and $U \cap W = \{\mathbf{0}\}$.

b. Let U be the xy-plane and W the z-axis in R^3. Is it true that $R^3 = U \oplus W$? Explain.

c. Let U be the xy-plane and W the yz-plane in R^3. Can every vector in R^3 be expressed as the sum of a vector in U and a vector in W? Is it true that $R^3 = U \oplus W$? Explain.

CHAPTER 5

Eigenvalues and Eigenvectors

CHAPTER CONTENTS

Introduction

In this chapter we will focus on classes of scalars and vectors known as "eigenvalues" and "eigenvectors," terms derived from the German word *eigen*, meaning "own," "peculiar to," "characteristic," or "individual." The underlying idea first appeared in the study of rotational motion but was later used to classify various kinds of surfaces and to describe solutions of certain differential equations. In the early 1900s it was applied to matrices and matrix transformations, and today it has applications in such diverse fields as computer graphics, mechanical vibrations, heat flow, population dynamics, quantum mechanics, and economics, to name just a few.

5.1 Eigenvalues and Eigenvectors

In this section we will define the notions of "eigenvalue" and "eigenvector" and discuss some of their basic properties.

Definition of Eigenvalue and Eigenvector

We begin with the main definition in this section.

Definition 1

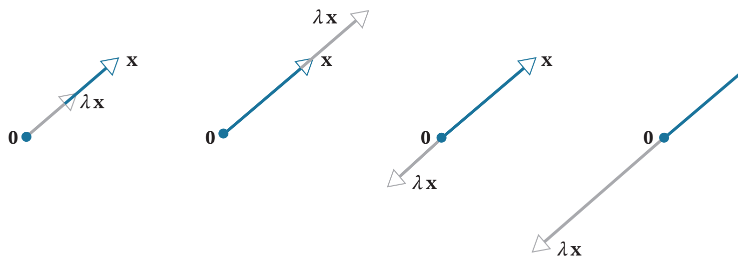
If A is an $n \times n$ matrix, then a nonzero vector $\mathbf{x}$ in R^n is called an ***eigenvector*** of A (or of the matrix operator T_A) if $A\mathbf{x}$ is a scalar multiple of $\mathbf{x}$; that is,

$$A\mathbf{x} = \lambda\mathbf{x}$$

for some scalar λ. The scalar λ is called an ***eigenvalue*** of A (or of T_A), and $\mathbf{x}$ is said to be an ***eigenvector corresponding to*** λ.

The requirement that an eigenvector be nonzero is imposed to avoid the unimportant case $A\mathbf{0} = \lambda\mathbf{0}$, which holds for every A and λ.

In general, the image of a vector $\mathbf{x}$ under multiplication by a square matrix A differs from $\mathbf{x}$ in both magnitude and direction. However, in the special case where $\mathbf{x}$ is an eigenvector of A, multiplication by A leaves the direction unchanged. For example, in R^2 or R^3 multiplication by A maps each eigenvector $\mathbf{x}$ of A (if any) along the same line through the origin as $\mathbf{x}$. Depending on the sign and magnitude of the eigenvalue λ corresponding to $\mathbf{x}$, the operation $A\mathbf{x} = \lambda\mathbf{x}$ compresses or stretches $\mathbf{x}$ by a factor of λ, with a reversal of direction in the case where λ is negative (**Figure 5.1.1**).



(*a*) $0 \le \lambda \le 1$ (*b*) $\lambda \ge 1$ (*c*) $-1 \le \lambda \le 0$ (*d*) $\lambda \le -1$

FIGURE 5.1.1

EXAMPLE 1 | Eigenvector of a 2 × 2 Matrix

The vector $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is an eigenvector of

$$A = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$$

corresponding to the eigenvalue $\lambda = 3$, since

$$A\mathbf{x} = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix} = 3\mathbf{x}$$

Geometrically, multiplication by A has stretched the vector $\mathbf{x}$ by a factor of 3 (**Figure 5.1.2**).

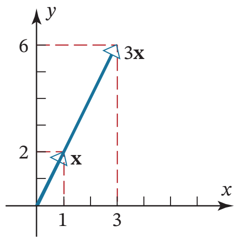


FIGURE 5.1.2

Computing Eigenvalues and Eigenvectors

Our next objective is to obtain a general procedure for finding eigenvalues and eigenvectors of an $n \times n$ matrix A. We will begin with the problem of finding the eigenvalues of A. Note first that the equation $A\mathbf{x} = \lambda\mathbf{x}$ can be rewritten as $A\mathbf{x} = \lambda I\mathbf{x}$, or equivalently as

$$(\lambda I - A)\mathbf{x} = \mathbf{0}$$

For λ to be an eigenvalue of A this equation must have a nonzero solution for $\mathbf{x}$. But it follows from parts (b) and (g) of Theorem 4.10.2 that this is so if and only if the coefficient matrix $\lambda I - A$ has a zero determinant. Thus, we have the following result.

Note that if $(A)_{ij} = a_{ij}$, then the left side of formula (1) can be written in expanded form as

$$\begin{vmatrix} \lambda - a_{11} & a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix}$$

Theorem 5.1.1

If A is an $n \times n$ matrix, then λ is an eigenvalue of A if and only if it satisfies the equation

$$\det(\lambda I - A) = 0 \tag{1}$$

This is called the ***characteristic equation*** of A.

EXAMPLE 2 | Finding Eigenvalues

In Example 1 we observed that $\lambda = 3$ is an eigenvalue of the matrix

$$A = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$$

but we did not explain how we found it. Use the characteristic equation to find all eigenvalues of this matrix.

Solution It follows from Formula (1) that the eigenvalues of A are the solutions of the equation $\det(\lambda I - A) = 0$, which we can write as

$$\begin{vmatrix} \lambda - 3 & 0 \\ -8 & \lambda + 1 \end{vmatrix} = 0$$

from which we obtain

$$(\lambda - 3)(\lambda + 1) = 0 \tag{2}$$

This shows that the eigenvalues of A are $\lambda = 3$ and $\lambda = -1$. Thus, in addition to the eigenvalue $\lambda = 3$ noted in Example 1, we have discovered a second eigenvalue $\lambda = -1$.

When the determinant $\det(\lambda I - A)$ in (1) is expanded, the characteristic equation of A takes the form

$$\boxed{\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0} \tag{3}$$

where the left side of this equation is a polynomial of degree n in which the coefficient of λ^n is 1 (Exercise 37). The polynomial

$$\boxed{p(\lambda) = \lambda^n + c_1\lambda^{n-1} + \cdots + c_n} \tag{4}$$

is called the ***characteristic polynomial*** of A. For example, it follows from (2) that the characteristic polynomial of the 2×2 matrix in Example 2 is

$$p(\lambda) = (\lambda - 3)(\lambda + 1) = \lambda^2 - 2\lambda - 3$$

which is a polynomial of degree 2.

Since a polynomial of degree n has at most n distinct roots, it follows from (3) that the characteristic equation of an $n \times n$ matrix A has at most n distinct solutions and consequently the matrix has at most n distinct eigenvalues. Since some of these solutions may be complex numbers, it is possible for a matrix to have complex eigenvalues, even if the matrix itself has real entries. We will discuss this issue in more detail later, but for now we will focus on examples in which the eigenvalues are real numbers.

EXAMPLE 3 | Eigenvalues of a 3 × 3 Matrix

Find the eigenvalues of

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{bmatrix}$$

Solution The characteristic polynomial of A is

$$\det(\lambda I - A) = \det\begin{bmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ -4 & 17 & \lambda - 8 \end{bmatrix} = \lambda^3 - 8\lambda^2 + 17\lambda - 4$$

The eigenvalues of A must therefore satisfy the cubic equation

$$\lambda^3 - 8\lambda^2 + 17\lambda - 4 = 0 \tag{5}$$

To solve this equation, we will begin by searching for integer solutions. This task can be simplified by exploiting the fact that all integer solutions (if there are any) of a polynomial equation with *integer coefficients*

$$\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0$$

must be divisors of the constant term, c_n. Thus, the only possible integer solutions of (5) are the divisors of -4, that is, $\pm 1, \pm 2, \pm 4$. Successively substituting these values in (5) shows that $\lambda = 4$ is an integer solution and hence that $\lambda - 4$ is a factor of the left side of (5). Dividing $\lambda - 4$ into $\lambda^3 - 8\lambda^2 + 17\lambda - 4$ shows that (5) can be rewritten as

$$(\lambda - 4)(\lambda^2 - 4\lambda + 1) = 0$$

Thus, the remaining solutions of (5) satisfy the quadratic equation

$$\lambda^2 - 4\lambda + 1 = 0$$

which can be solved by the quadratic formula. Thus, the eigenvalues of A are

$$\lambda = 4, \quad \lambda = 2 + \sqrt{3}, \quad \text{and} \quad \lambda = 2 - \sqrt{3}$$

In applications involving large matrices it is often not feasible to compute the characteristic equation directly, so other methods must be used to find eigenvalues. We will consider such methods in Chapter 9.

EXAMPLE 4 | Eigenvalues of an Upper Triangular Matrix

Find the eigenvalues of the upper triangular matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}$$

Solution Recalling that the determinant of a triangular matrix is the product of the entries on the main diagonal (Theorem 2.1.2), we obtain

$$\det(\lambda I - A) = \det\begin{bmatrix} \lambda - a_{11} & -a_{12} & -a_{13} & -a_{14} \\ 0 & \lambda - a_{22} & -a_{23} & -a_{24} \\ 0 & 0 & \lambda - a_{33} & -a_{34} \\ 0 & 0 & 0 & \lambda - a_{44} \end{bmatrix}$$
$$= (\lambda - a_{11})(\lambda - a_{22})(\lambda - a_{33})(\lambda - a_{44})$$

Thus, the characteristic equation is

$$(\lambda - a_{11})(\lambda - a_{22})(\lambda - a_{33})(\lambda - a_{44}) = 0$$

and the eigenvalues are

$$\lambda = a_{11}, \quad \lambda = a_{22}, \quad \lambda = a_{33}, \quad \lambda = a_{44}$$

which are precisely the diagonal entries of A.

The following general theorem should be evident from the computations in the preceding example.

Theorem 5.1.2

If A is an $n \times n$ triangular matrix (upper triangular, lower triangular, or diagonal), then the eigenvalues of A are the entries on the main diagonal of A.

Had Theorem 5.1.2 been available earlier, we could have anticipated the result obtained in Example 2.

EXAMPLE 5 | Eigenvalues of a Lower Triangular Matrix

By inspection, the eigenvalues of the lower triangular matrix

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

are $\lambda = \frac{1}{2}$, $\lambda = \frac{2}{3}$, and $\lambda = -\frac{1}{4}$.

The following theorem gives some alternative ways of describing eigenvalues.

Theorem 5.1.3

If A is an $n \times n$ matrix, the following statements are equivalent.

(*a*) λ is an eigenvalue of A.

(*b*) λ is a solution of the characteristic equation $\det(\lambda I - A) = 0$.

(*c*) The system of equations $(\lambda I - A)\mathbf{x} = \mathbf{0}$ has nontrivial solutions.

(*d*) There is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$.

Finding Eigenvectors and Bases for Eigenspaces

Now that we know how to find the eigenvalues of a matrix, we will consider the problem of finding the corresponding eigenvectors. By definition, the eigenvectors of A corresponding to an eigenvalue λ are the nonzero vectors that satisfy

$$(\lambda I - A)\mathbf{x} = \mathbf{0}$$

Thus, we can find the eigenvectors of A corresponding to λ by finding the nonzero vectors in the solution space of this linear system. This solution space, which is called the ***eigenspace*** of A corresponding to λ, can also be viewed as:

1. the null space of the matrix $\lambda I - A$
2. the kernel of the matrix operator $T_{\lambda I - A}: R^n \to R^n$
3. the set of vectors for which $A\mathbf{x} = \lambda\mathbf{x}$

Notice that $\mathbf{x} = \mathbf{0}$ is in every eigenspace but is not an eigenvector (see Definition 1). In the exercises we will ask you to show that this is the *only* vector that distinct eigenspaces have in common.

EXAMPLE 6 | Bases for Eigenspaces

Find bases for the eigenspaces of the matrix

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

Solution The characteristic equation of A is

$$\begin{vmatrix} \lambda + 1 & -3 \\ -2 & \lambda \end{vmatrix} = \lambda(\lambda + 1) - 6 = (\lambda - 2)(\lambda + 3) = 0$$

so the eigenvalues of A are $\lambda = 2$ and $\lambda = -3$. Thus, there are two eigenspaces of A, one for each eigenvalue. By definition,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

is an eigenvector of A corresponding to an eigenvalue λ if and only if $(\lambda I - A)\mathbf{x} = \mathbf{0}$, that is,

$$\begin{bmatrix} \lambda+1 & -3 \\ -2 & \lambda \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

In the case where $\lambda = 2$ this equation becomes

$$\begin{bmatrix} 3 & -3 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

whose general solution is

$$x_1 = t, \quad x_2 = t$$

(verify). Since this can be written in matrix form as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} t \\ t \end{bmatrix} = t\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

it follows that

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

is a basis for the eigenspace corresponding to $\lambda = 2$. We leave it for you to follow the pattern of these computations and show that

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

is a basis for the eigenspace corresponding to $\lambda = -3$.

Figure 5.1.3 illustrates the geometric effect of multiplication by the matrix A in Example 6. The eigenspace corresponding to $\lambda = 2$ is the line L_1 through the origin and the point $(1, 1)$, and the eigenspace corresponding to $\lambda = 3$ is the line L_2 through the origin and the point $\left(-\frac{3}{2}, 1\right)$. As indicated in the figure, multiplication by A maps each vector in L_1 back into L_1, scaling it by a factor of 2, and it maps each vector in L_2 back into L_2, scaling it by a factor of -3.

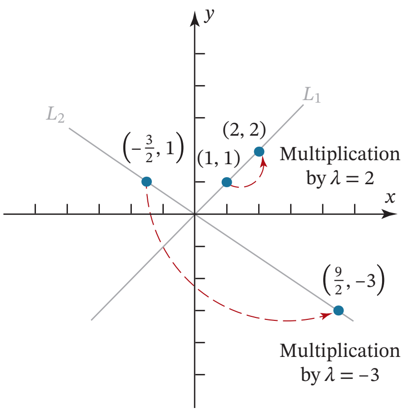
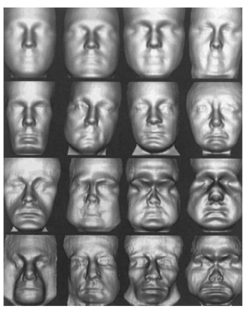


FIGURE 5.1.3

Historical Note



Methods of linear algebra are used in the emerging field of computerized face recognition. Researchers are working with the idea that every human face in a racial group is a combination of a few dozen primary shapes. For example, by analyzing three-dimensional scans of many faces, researchers at Rockefeller University have produced both an average head shape in the Caucasian group—dubbed the ***meanhead*** (top row left in the figure to the left)—and a set of standardized variations from that shape, called ***eigenheads*** (15 of which are shown in the picture). These are so named because they are eigenvectors of a certain matrix that stores digitized facial information. Face shapes are represented mathematically as linear combinations of the eigenheads.

[Image: © Dr. Joseph J. Atick, adapted from *Scientific American*]

EXAMPLE 7 | Eigenvectors and Bases for Eigenspaces

Find bases for the eigenspaces of

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

Solution The characteristic equation of A is $\lambda^3 - 5\lambda^2 + 8\lambda - 4 = 0$, or in factored form, $(\lambda - 1)(\lambda - 2)^2 = 0$ (verify). Thus, the distinct eigenvalues of A are $\lambda = 1$ and $\lambda = 2$, so there are two eigenspaces of A.

By definition,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

is an eigenvector of A corresponding to λ if and only if $\mathbf{x}$ is a nontrivial solution of $(\lambda I - A)\mathbf{x} = \mathbf{0}$, or in matrix form,

$$\begin{bmatrix} \lambda & 0 & 2 \\ -1 & \lambda - 2 & -1 \\ -1 & 0 & \lambda - 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{6}$$

In the case where $\lambda = 2$, Formula (6) becomes

$$\begin{bmatrix} 2 & 0 & 2 \\ -1 & 0 & -1 \\ -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Solving this system using Gaussian elimination yields (verify)

$$x_1 = -s, \quad x_2 = t, \quad x_3 = s$$

Thus, the eigenvectors of A corresponding to $\lambda = 2$ are the nonzero vectors of the form

$$\mathbf{x} = \begin{bmatrix} -s \\ t \\ s \end{bmatrix} = \begin{bmatrix} -s \\ 0 \\ s \end{bmatrix} + \begin{bmatrix} 0 \\ t \\ 0 \end{bmatrix} = s\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + t\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Since

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

are linearly independent (why?), these vectors form a basis for the eigenspace corresponding to $\lambda = 2$.

If $\lambda = 1$, then (6) becomes

$$\begin{bmatrix} 1 & 0 & 2 \\ -1 & -1 & -1 \\ -1 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Solving this system yields (verify)

$$x_1 = -2s, \quad x_2 = s, \quad x_3 = s$$

Thus, the eigenvectors corresponding to $\lambda = 1$ are the nonzero vectors of the form

$$\begin{bmatrix} -2s \\ s \\ s \end{bmatrix} = s\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \quad \text{so that} \quad \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

is a basis for the eigenspace corresponding to $\lambda = 1$.

Eigenvalues and Invertibility

The next theorem establishes a relationship between the eigenvalues and the invertibility of a matrix.

Theorem 5.1.4

A square matrix A is invertible if and only if $\lambda = 0$ is not an eigenvalue of A.

Proof Assume that A is an $n \times n$ matrix and observe first that $\lambda = 0$ is a solution of the characteristic equation

$$\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0$$

if and only if the constant term c_n is zero. Thus, it suffices to prove that A is invertible if and only if $c_n \neq 0$. But

$$\det(\lambda I - A) = \lambda^n + c_1\lambda^{n-1} + \cdots + c_n$$

or, on setting $\lambda = 0$,

$$\det(-A) = c_n \quad \text{or} \quad (-1)^n \det(A) = c_n$$

It follows from the last equation that $\det(A) = 0$ if and only if $c_n = 0$, and this in turn implies that A is invertible if and only if $c_n \neq 0$. ■

EXAMPLE 8 | Eigenvalues and Invertibility

The matrix A in Example 7 is invertible since it has eigenvalues $\lambda = 1$ and $\lambda = 2$, neither of which is zero. We leave it for you to check this conclusion by showing that $\det(A) \neq 0$.

More on the Equivalence Theorem

As our final result in this section, we will use Theorem 5.1.4 to add one additional part to Theorem 4.9.8.

Theorem 5.1.5

Equivalent Statements

If A is an $n \times n$ matrix in which there are no duplicate rows and no duplicate columns, then the following statements are equivalent.

(*a*) A is invertible.
(*b*) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
(*c*) The reduced row echelon form of A is I_n.
(*d*) A is expressible as a product of elementary matrices.
(*e*) $A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.
(*f*) $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.
(*g*) $\det(A) \neq 0$.
(*h*) The column vectors of A are linearly independent.
(*i*) The row vectors of A are linearly independent.
(*j*) The column vectors of A span R^n.
(*k*) The row vectors of A span R^n.
(*l*) The column vectors of A form a basis for R^n.
(*m*) The row vectors of A form a basis for R^n.
(*n*) A has rank n.
(*o*) A has nullity 0.
(*p*) The orthogonal complement of the null space of A is R^n.
(*q*) The orthogonal complement of the row space of A is $\{\mathbf{0}\}$.
(*r*) $\lambda = 0$ is not an eigenvalue of A.

Exercise Set 5.1

In Exercises 1–4, confirm by multiplication that $\mathbf{x}$ *is an eigenvector of* A*, and find the corresponding eigenvalue.*

1. $A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

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

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

4. $A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

In each part of Exercises 5–6, find the characteristic equation, the eigenvalues, and bases for the eigenspaces of the matrix.

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

c. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ **d.** $\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$

6. a. $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ **b.** $\begin{bmatrix} 2 & -3 \\ 0 & 2 \end{bmatrix}$

c. $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ **d.** $\begin{bmatrix} 1 & 2 \\ -2 & -1 \end{bmatrix}$

In Exercises 7–12, find the characteristic equation, the eigenvalues, and bases for the eigenspaces of the matrix.

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

9. $\begin{bmatrix} 6 & 3 & -8 \\ 0 & -2 & 0 \\ 1 & 0 & -3 \end{bmatrix}$ **10.** $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

11. $\begin{bmatrix} 4 & 0 & -1 \\ 0 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$ **12.** $\begin{bmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{bmatrix}$

In Exercises 13–14, find the characteristic equation of the matrix by inspection.

13. $\begin{bmatrix} 3 & 0 & 0 \\ -2 & 7 & 0 \\ 4 & 8 & 1 \end{bmatrix}$ **14.** $\begin{bmatrix} 9 & -8 & 6 & 3 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix}$

In Exercises 15–16, find the eigenvalues and a basis for each eigenspace of the linear operator defined by the stated formula. [Suggestion: *Work with the standard matrix for the operator.*]

15. $T(x, y) = (x + 4y, 2x + 3y)$

16. $T(x, y, z) = (2x - y - z,\ x - z,\ -x + y + 2z)$

17. **(*Calculus required*)** Let $D^2\colon C^\infty(-\infty, \infty) \to C^\infty(-\infty, \infty)$ be the operator that maps a function into its second derivative.

a. Show that D^2 is linear.

b. Show that if ω is a positive constant, then $\sin\sqrt{\omega}x$ and $\cos\sqrt{\omega}x$ are eigenvectors of D^2, and find their corresponding eigenvalues.

18. **(*Calculus required*)** Let $D^2\colon C^\infty \to C^\infty$ be the linear operator in Exercise 17. Show that if ω is a positive constant, then $\sinh\sqrt{\omega}x$ and $\cosh\sqrt{\omega}x$ are eigenvectors of D^2, and find their corresponding eigenvalues.

In each part of Exercises 19–20, find the eigenvalues and the corresponding eigenspaces of the stated matrix operator on R^2*. Use geometric reasoning to find the answers. No computations are needed.*

19. a. Reflection about the line $y = x$.

b. Orthogonal projection onto the x-axis.

c. Rotation about the origin through a positive angle of 90°.

d. Contraction with factor k $(0 \le k < 1)$.

e. Shear in the x-direction by a factor k $(k \neq 0)$.

20. a. Reflection about the y-axis.

b. Rotation about the origin through a positive angle of 180°.

c. Dilation with factor k $(k > 1)$.

d. Expansion in the y-direction with factor k $(k > 1)$.

e. Shear in the y-direction by a factor k $(k \neq 0)$.

In each part of Exercises 21–22, find the eigenvalues and the corresponding eigenspaces of the stated matrix operator on R^3*. Use geometric reasoning to find the answers. No computations are needed.*

21. a. Reflection about the xy-plane.

b. Orthogonal projection onto the xz-plane.

c. Counterclockwise rotation about the positive x-axis through an angle of 90°.

d. Contraction with factor k $(0 \le k < 1)$.

22. a. Reflection about the xz-plane.

b. Orthogonal projection onto the yz-plane.

c. Counterclockwise rotation about the positive y-axis through an angle of 180°.

d. Dilation with factor k $(k > 1)$.

23. Let A be a 2×2 matrix, and call a line through the origin of R^2 ***invariant*** under A if $A\mathbf{x}$ lies on the line when $\mathbf{x}$ does. Find equations for all lines in R^2, if any, that are invariant under the given matrix.

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

24. Find $\det(A)$ given that A has $p(\lambda)$ as its characteristic polynomial.

a. $p(\lambda) = \lambda^3 - 2\lambda^2 + \lambda + 5$

b. $p(\lambda) = \lambda^4 - \lambda^3 + 7$

[*Hint:* See the proof of Theorem 5.1.4.]

25. Suppose that the characteristic polynomial of some matrix A is found to be $p(\lambda) = (\lambda - 1)(\lambda - 3)^2(\lambda - 4)^3$. In each part, answer the question and explain your reasoning.

a. What is the size of A?

b. Is A invertible?

c. How many eigenspaces does A have?

26. The eigenvectors that we have been studying are sometimes called ***right eigenvectors*** to distinguish them from ***left eigenvectors***, which are $n \times 1$ column matrices $\mathbf{x}$ that satisfy the equation $\mathbf{x}^T A = \mu \mathbf{x}^T$ for some scalar μ. For a given matrix A, how are the right eigenvectors and their corresponding eigenvalues related to the left eigenvectors and their corresponding eigenvalues?

27. Find a 3×3 matrix A that has eigenvalues 1, -1, and 0, and for which

$$\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

are their corresponding eigenvectors.

Working with Proofs

28. Prove that the characteristic equation of a 2×2 matrix A can be expressed as $\lambda^2 - \operatorname{tr}(A)\lambda + \det(A) = 0$, where $\operatorname{tr}(A)$ is the trace of A.

29. Use the result in Exercise 28 to show that if

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

then the solutions of the characteristic equation of A are

$$\lambda = \tfrac{1}{2}\left[(a+d) \pm \sqrt{(a-d)^2 + 4bc}\,\right]$$

Use this result to show that A has

a. two distinct real eigenvalues if $(a-d)^2 + 4bc > 0$.

b. two repeated real eigenvalues if $(a-d)^2 + 4bc = 0$.

c. complex conjugate eigenvalues if $(a-d)^2 + 4bc < 0$.

30. Let A be the matrix in Exercise 29. Show that if $b \neq 0$, then

$$\mathbf{x}_1 = \begin{bmatrix} -b \\ a - \lambda_1 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} -b \\ a - \lambda_2 \end{bmatrix}$$

are eigenvectors of A that correspond, respectively, to the eigenvalues

$$\lambda_1 = \tfrac{1}{2}\left[(a+d) + \sqrt{(a-d)^2 + 4bc}\,\right]$$

and

$$\lambda_2 = \tfrac{1}{2}\left[(a+d) - \sqrt{(a-d)^2 + 4bc}\,\right]$$

31. Use the result of Exercise 28 to prove that if

$$p(\lambda) = \lambda^2 + c_1\lambda + c_2$$

is the characteristic polynomial of a 2×2 matrix, then

$$p(A) = A^2 + c_1 A + c_2 I = 0$$

(Stated informally, A satisfies its characteristic equation. This result is true as well for $n \times n$ matrices.)

32. Prove: If a, b, c, and d are integers such that $a + b = c + d$, then

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

has integer eigenvalues.

33. Prove: If λ is an eigenvalue of an invertible matrix A and $\mathbf{x}$ is a corresponding eigenvector, then $1/\lambda$ is an eigenvalue of A^{-1} and $\mathbf{x}$ is a corresponding eigenvector.

34. Prove: If λ is an eigenvalue of A, $\mathbf{x}$ is a corresponding eigenvector, and s is a scalar, then $\lambda - s$ is an eigenvalue of $A - sI$ and $\mathbf{x}$ is a corresponding eigenvector.

35. Prove: If λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector, then $s\lambda$ is an eigenvalue of sA for every scalar s and $\mathbf{x}$ is a corresponding eigenvector.

36. Find the eigenvalues and bases for the eigenspaces of

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

and then use Exercises 33 and 34 to find the eigenvalues and bases for the eigenspaces of

a. A^{-1} **b.** $A - 3I$ **c.** $A + 2I$

37. Prove that the characteristic polynomial of an $n \times n$ matrix A has degree n and that the coefficient of λ^n in that polynomial is 1.

38. **a.** Prove that if A is a square matrix, then A and A^T have the same eigenvalues. [*Hint:* Look at the characteristic equation $\det(\lambda I - A) = 0$.]

b. Show that A and A^T need not have the same eigenspaces. [*Hint:* Use the result in Exercise 30 to find a 2×2 matrix for which A and A^T have different eigenspaces.]

39. Prove that the intersection of any two distinct eigenspaces of a matrix A is $\{\mathbf{0}\}$.

True-False Exercises

TF. In parts **(a)–(f)** determine whether the statement is true or false, and justify your answer.

a. If A is a square matrix and $A\mathbf{x} = \lambda\mathbf{x}$ for some nonzero scalar λ, then $\mathbf{x}$ is an eigenvector of A.

b. If λ is an eigenvalue of a matrix A, then the linear system $(\lambda I - A)\mathbf{x} = \mathbf{0}$ has only the trivial solution.

c. If the characteristic polynomial of a matrix A is

$$p(\lambda) = \lambda^2 + 1$$

then A is invertible.

d. If λ is an eigenvalue of a matrix A, then the eigenspace of A corresponding to λ is the set of eigenvectors of A corresponding to λ.

e. The eigenvalues of a matrix A are the same as the eigenvalues of the reduced row echelon form of A.

f. If 0 is an eigenvalue of a matrix A, then the set of columns of A is linearly independent.

Working with Technology

T1. For the given matrix A, find the characteristic polynomial and the eigenvalues, and then use the method of Example 7 to find bases for the eigenspaces.

$$A = \begin{bmatrix} -8 & 33 & 38 & 173 & -30 \\ 0 & 0 & -1 & -4 & 0 \\ 0 & 0 & -5 & -25 & 1 \\ 0 & 0 & 1 & 5 & 0 \\ 4 & -16 & -19 & -86 & 15 \end{bmatrix}$$

T2. The Cayley–Hamilton Theorem states that every square matrix satisfies its characteristic equation; that is, if A is an $n \times n$ matrix whose characteristic equation is

$$\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0$$

then $A^n + c_1A^{n-1} + \cdots + c_n = 0$.

a. Verify the Cayley–Hamilton Theorem for the matrix

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

b. Use the result in Exercise 28 to prove the Cayley–Hamilton Theorem for 2×2 matrices.

5.2 Diagonalization

In this section we will be concerned with the problem of finding a basis for R^n that consists of eigenvectors of an $n \times n$ matrix A. Such bases can be used to study geometric properties of A and to simplify various numerical computations. These bases are also of physical significance in a wide variety of applications, some of which will be considered later in this text.

The Matrix Diagonalization Problem

Products of the form $P^{-1}AP$ in which A and P are $n \times n$ matrices and P is invertible will be our main topic of study in this section. There are various ways to think about such products, one of which is to view them as transformations of the form

$$A \to P^{-1}AP$$

in which the matrix A is mapped into the matrix $P^{-1}AP$. These are called ***similarity transformations***. Such transformations are important because they preserve many properties of the matrix A. For example, if we let $B = P^{-1}AP$, then A and B have the same determinant since

$$\begin{aligned} \det(B) &= \det(P^{-1}AP) = \det(P^{-1})\det(A)\det(P) \\ &= \frac{1}{\det(P)}\det(A)\det(P) = \det(A) \end{aligned}$$

In general, any property that is preserved by a similarity transformation is called a ***similarity invariant*** and is said to be ***invariant under similarity***. **Table 1** lists the most important similarity invariants. The proofs of some of these are given as exercises.

TABLE 1 Similarity Invariants

Property	Description
Determinant	A and $P^{-1}AP$ have the same determinant.
Invertibility	A is invertible if and only if $P^{-1}AP$ is invertible.
Rank	A and $P^{-1}AP$ have the same rank.
Nullity	A and $P^{-1}AP$ have the same nullity.
Trace	A and $P^{-1}AP$ have the same trace.
Characteristic polynomial	A and $P^{-1}AP$ have the same characteristic polynomial.
Eigenvalues	A and $P^{-1}AP$ have the same eigenvalues.
Eigenspace dimension	If λ is an eigenvalue of A (and hence of $P^{-1}AP$) then the eigenspace of A corresponding to λ and the eigenspace of $P^{-1}AP$ corresponding to λ have the same dimension.

We will find the following terminology useful in our study of similarity transformations.

Definition 1

If A and B are square matrices, then we say that ***B is similar to A*** if there is an invertible matrix P such that $B = P^{-1}AP$.

Note that if B is similar to A, then it is also true that A is similar to B since we can express A as $A = Q^{-1}BQ$ by taking $Q = P^{-1}$. This being the case, we will usually say that A and B are ***similar matrices*** if either is similar to the other.

Because diagonal matrices have such a simple form, it is natural to inquire whether a given $n \times n$ matrix A is similar to a matrix of this type. Should this turn out to be the case, and should we be able to actually find a diagonal matrix D that is similar to A, then we would be able to ascertain many of the similarity invariant properties of A directly from the diagonal entries of D. For example, the diagonal entries of D will be the eigenvalues of A (Theorem 5.1.2), and the product of the diagonal entries of D will be the determinant of A (Theorem 2.1.2). This leads us to introduce the following terminology.

Definition 2

A square matrix A is said to be ***diagonalizable*** if it is similar to some diagonal matrix; that is, if there exists an invertible matrix P such that $P^{-1}AP$ is diagonal. In this case the matrix P is said to ***diagonalize*** A.

The following theorem and the ideas used in its proof will provide us with a roadmap for devising a technique for determining whether a matrix is diagonalizable and, if so, for finding a matrix P that will perform the diagonalization.

Part (b) of Theorem 5.2.1 is equivalent to saying that there is a *basis* for R^n consisting of eigenvectors of A. Why?

Theorem 5.2.1

If A is an $n \times n$ matrix, the following statements are equivalent.

(a) A is diagonalizable.

(b) A has n linearly independent eigenvectors.

Proof (a) ⇒ (b) Since A is assumed to be diagonalizable, it follows that there exist an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$ or, equivalently,

$$AP = PD \tag{1}$$

If we denote the column vectors of P by $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$, and if we assume that the diagonal entries of D are $\lambda_1, \lambda_2, \ldots, \lambda_n$, then by Formula (6) of Section 1.3 the left side of (1) can be expressed as

$$AP = A[\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_n] = [A\mathbf{p}_1 \quad A\mathbf{p}_2 \quad \cdots \quad A\mathbf{p}_n]$$

and, as noted in the comment following Example 1 of Section 1.7, the right side of (1) can be expressed as

$$PD = [\lambda_1\mathbf{p}_1 \quad \lambda_2\mathbf{p}_2 \quad \cdots \quad \lambda_n\mathbf{p}_n]$$

Thus, it follows from (1) that

$$A\mathbf{p}_1 = \lambda_1\mathbf{p}_1, \quad A\mathbf{p}_2 = \lambda_2\mathbf{p}_2, \ldots, \quad A\mathbf{p}_n = \lambda_n\mathbf{p}_n \tag{2}$$

Since P is invertible, we know from Theorem 5.1.5 that its column vectors $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ are linearly independent (and hence nonzero). Thus, it follows from (2) that these n column vectors are eigenvectors of A. ■

Proof (b) ⇒ (a) Assume that A has n linearly independent eigenvectors, $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$, and that $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the corresponding eigenvalues. If we let

$$P = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_n]$$

and if we let D be the diagonal matrix that has $\lambda_1, \lambda_2, \ldots, \lambda_n$ as its successive diagonal entries, then

$$\begin{aligned} AP &= A[\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_n] = [A\mathbf{p}_1 \quad A\mathbf{p}_2 \quad \cdots \quad A\mathbf{p}_n] \\ &= [\lambda_1\mathbf{p}_1 \quad \lambda_2\mathbf{p}_2 \quad \cdots \quad \lambda_n\mathbf{p}_n] = PD \end{aligned}$$

Since the column vectors of P are linearly independent, it follows from Theorem 5.1.5 that P is invertible, so that this last equation can be rewritten as $P^{-1}AP = D$, which shows that A is diagonalizable. ■

Whereas Theorem 5.2.1 tells us that we need to find n linearly independent eigenvectors to diagonalize a matrix, the following theorem tells us where such vectors might be found. Part (a) is proved at the end of this section, and part (b) is an immediate consequence of part (a) and Theorem 5.2.1 (why?).

Theorem 5.2.2

(a) If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of a matrix A, and if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are corresponding eigenvectors, then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a linearly independent set.

(b) An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

Remark Part (a) of Theorem 5.2.2 is a special case of a more general result: Specifically, if $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues, and if $S_1, S_2, \ldots, S_k$ are corresponding sets of linearly independent eigenvectors, then the *union* of these sets is linearly independent.

Procedure for Diagonalizing a Matrix

Theorem 5.2.1 guarantees that an $n \times n$ matrix A with n linearly independent eigenvectors is diagonalizable, and the proof of that theorem together with Theorem 5.2.2 suggests the following procedure for diagonalizing A.

A Procedure for Diagonalizing an $n \times n$ Matrix

Step 1. Determine first whether the matrix is actually diagonalizable by searching for n linearly independent eigenvectors. One way to do this is to find a basis for each eigenspace and count the total number of vectors obtained. If there is a total of n vectors, then the matrix is diagonalizable, and if the total is less than n, then it is not.

Step 2. If you ascertained that the matrix is diagonalizable, then form the matrix $P = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_n]$ whose column vectors are the n basis vectors you obtained in Step 1.

Step 3. $P^{-1}AP$ will be a diagonal matrix whose successive diagonal entries are the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ that correspond to the successive columns of P.

EXAMPLE 1 | Finding a Matrix P That Diagonalizes a Matrix A

Find a matrix P that diagonalizes

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

Solution In Example 7 of the preceding section we found the characteristic equation of A to be

$$(\lambda - 1)(\lambda - 2)^2 = 0$$

and we found the following bases for the eigenspaces:

$$\lambda = 2: \quad \mathbf{p}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{p}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}; \qquad \lambda = 1: \quad \mathbf{p}_3 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

There are three basis vectors in total, so the matrix

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

diagonalizes A. As a check, you should verify that

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

In general, there is no preferred order for the columns of P. Since the ith diagonal entry of $P^{-1}AP$ is an eigenvalue for the ith column vector of P, changing the order of the columns of P just changes the order of the eigenvalues on the diagonal of $P^{-1}AP$. Thus, had we written

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

in the preceding example, we would have obtained

$$P^{-1}AP = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

EXAMPLE 2 | A Matrix That Is Not Diagonalizable

Show that the following matrix is not diagonalizable:

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

Solution The characteristic polynomial of A is

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - 1 & 0 & 0 \\ -1 & \lambda - 2 & 0 \\ 3 & -5 & \lambda - 2 \end{vmatrix} = (\lambda - 1)(\lambda - 2)^2$$

so the characteristic equation is

$$(\lambda - 1)(\lambda - 2)^2 = 0$$

and the distinct eigenvalues of A are $\lambda = 1$ and $\lambda = 2$. We leave it for you to show that bases for the eigenspaces are

$$\lambda = 1: \quad \mathbf{p}_1 = \begin{bmatrix} \frac{1}{8} \\ -\frac{1}{8} \\ 1 \end{bmatrix}; \qquad \lambda = 2: \quad \mathbf{p}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Since A is a 3×3 matrix and there are only two basis vectors in total, A is not diagonalizable.

Alternative Solution If you are concerned only in determining whether a matrix is diagonalizable and not with actually finding a diagonalizing matrix P, then it is not necessary to compute bases for the eigenspaces—it suffices to find the dimensions of the eigenspaces. For this example, the eigenspace corresponding to $\lambda = 1$ is the solution space of the system

$$\begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & 0 \\ 3 & -5 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Since the coefficient matrix has rank 2 (verify), the nullity of this matrix is 1 by Theorem 4.9.2, and hence the eigenspace corresponding to $\lambda = 1$ is one-dimensional.

The eigenspace corresponding to $\lambda = 2$ is the solution space of the system

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 3 & -5 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This coefficient matrix also has rank 2 and nullity 1 (verify), so the eigenspace corresponding to $\lambda = 2$ is also one-dimensional. Since the eigenspaces produce a total of two basis vectors, and since three are needed, the matrix A is not diagonalizable.

EXAMPLE 3 | Recognizing Diagonalizability

We saw in Example 3 of the preceding section that

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{bmatrix}$$

has three distinct eigenvalues: $\lambda = 4$, $\lambda = 2 + \sqrt{3}$, and $\lambda = 2 - \sqrt{3}$. Therefore, A is diagonalizable and

$$P^{-1}AP = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2+\sqrt{3} & 0 \\ 0 & 0 & 2-\sqrt{3} \end{bmatrix}$$

for some invertible matrix P. If needed, the matrix P can be found using the method shown in Example 1 of this section.

EXAMPLE 4 | Diagonalizability of Triangular Matrices

From Theorem 5.1.2, the eigenvalues of a triangular matrix are the entries on its main diagonal. Thus, a triangular matrix with distinct entries on the main diagonal is diagonalizable. For example,

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

is a diagonalizable matrix with eigenvalues $\lambda_1 = -1, \lambda_2 = 3, \lambda_3 = 5, \lambda_4 = -2$.

Eigenvalues of Powers of a Matrix

Since there are many applications in which it is necessary to compute high powers of a square matrix A, we will now turn our attention to that important problem. As we will see, the most efficient way to compute A^k, particularly for large values of k, is to first diagonalize A. But because diagonalizing a matrix A involves finding its eigenvalues and eigenvectors, we will need to know how these quantities are related to those of A^k. As an illustration, suppose that λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector. Then

$$A^2\mathbf{x} = A(A\mathbf{x}) = A(\lambda\mathbf{x}) = \lambda(A\mathbf{x}) = \lambda(\lambda\mathbf{x}) = \lambda^2\mathbf{x}$$

which shows not only that λ^2 is a eigenvalue of A^2 but that $\mathbf{x}$ is a corresponding eigenvector. In general, we have the following result.

Theorem 5.2.3

If k is a positive integer, λ is an eigenvalue of a matrix A, and $\mathbf{x}$ is a corresponding eigenvector, then λ^k is an eigenvalue of A^k and $\mathbf{x}$ is a corresponding eigenvector.

Note that diagonalizability is not a requirement in Theorem 5.2.3.

EXAMPLE 5 | Eigenvalues and Eigenvectors of Matrix Powers

In Example 2 we found the eigenvalues and corresponding eigenvectors of the matrix

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

Do the same for A^7.

Solution We know from Example 2 that the eigenvalues of A are $\lambda = 1$ and $\lambda = 2$, so the eigenvalues of A^7 are $\lambda = 1^7 = 1$ and $\lambda = 2^7 = 128$. The eigenvectors $\mathbf{p}_1$ and $\mathbf{p}_2$ obtained in Example 1 corresponding to the eigenvalues $\lambda = 1$ and $\lambda = 2$ of A are also the eigenvectors corresponding to the eigenvalues $\lambda = 1$ and $\lambda = 128$ of A^7.

Computing Powers of a Matrix

The problem of computing powers of a matrix is greatly simplified when the matrix is diagonalizable. To see why this is so, suppose that A is a diagonalizable $n \times n$ matrix, that P diagonalizes A, and that

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = D$$

Squaring both sides of this equation yields

$$(P^{-1}AP)^2 = \begin{bmatrix} \lambda_1^2 & 0 & \cdots & 0 \\ 0 & \lambda_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n^2 \end{bmatrix} = D^2$$

We can rewrite the left side of this equation as

$$(P^{-1}AP)^2 = P^{-1}APP^{-1}AP = P^{-1}AIAP = P^{-1}A^2P$$

from which we obtain the relationship $P^{-1}A^2P = D^2$. More generally, if k is a positive integer, then a similar computation will show that

$$P^{-1}A^kP = D^k = \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{bmatrix}$$

Formula (3) reveals that raising a diagonalizable matrix A to a positive integer power has the effect of raising its eigenvalues to that power.

which we can rewrite as

$$A^k = PD^kP^{-1} = P\begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{bmatrix}P^{-1} \tag{3}$$

EXAMPLE 6 | Powers of a Matrix

Use (3) to find A^{13}, where

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

Solution We showed in Example 1 that the matrix A is diagonalized by

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

and that

$$D = P^{-1}AP = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Thus, it follows from (3) that

$$A^{13} = PD^{13}P^{-1} = \begin{bmatrix} -1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2^{13} & 0 & 0 \\ 0 & 2^{13} & 0 \\ 0 & 0 & 1^{13} \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ -1 & 0 & -1 \end{bmatrix} \tag{4}$$

$$= \begin{bmatrix} -8190 & 0 & -16382 \\ 8191 & 8192 & 8191 \\ 8191 & 0 & 16383 \end{bmatrix}$$

Remark With the method in the preceding example, most of the work is in diagonalizing A. Once that work is done, it can be used to compute any power of A. Thus, to compute A^{1000} we need only change the exponents from 13 to 1000 in (4).

Geometric and Algebraic Multiplicity

Theorem 5.2.2(b) does not completely settle the diagonalizability question since it only guarantees that a square matrix with n distinct eigenvalues is diagonalizable; it does not preclude the possibility that there may exist diagonalizable matrices with fewer than n distinct eigenvalues. The following example shows that this is indeed the case.

EXAMPLE 7 | The Converse of Theorem 5.2.2(b) Is False

Consider the matrices

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad J = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

It follows from Theorem 5.1.2 that both of these matrices have only one distinct eigenvalue, namely $\lambda = 1$, and hence only one eigenspace. We leave it as an exercise for you to solve the characteristic equations

$$(\lambda I - I)\mathbf{x} = \mathbf{0} \quad \text{and} \quad (\lambda I - J)\mathbf{x} = \mathbf{0}$$

with $\lambda = 1$ and show that for I the eigenspace is three-dimensional (all of R^3) and for J it is one-dimensional, consisting of all scalar multiples of

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

This shows that the converse of Theorem 5.2.2(b) is false, since we have produced two 3×3 matrices with fewer than three distinct eigenvalues, one of which is diagonalizable and the other of which is not.

A full excursion into the study of diagonalizability is left for more advanced courses, but we will touch on one theorem that is important for a fuller understanding of this topic. It can be proved that if λ_0 is an eigenvalue of A, then the dimension of the eigenspace corresponding to λ_0 cannot exceed the number of times that $\lambda - \lambda_0$ appears as a factor of the characteristic polynomial of A. For example, in Examples 1 and 2 the characteristic polynomial is

$$(\lambda - 1)(\lambda - 2)^2$$

Thus, the eigenspace corresponding to $\lambda = 1$ is at most (hence exactly) one-dimensional, and the eigenspace corresponding to $\lambda = 2$ is at most two-dimensional. In Example 1 the eigenspace corresponding to $\lambda = 2$ actually had dimension 2, resulting in diagonalizability, but in Example 2 the eigenspace corresponding to $\lambda = 2$ had only dimension 1, resulting in nondiagonalizability.

There is some terminology that is related to these ideas. If λ_0 is an eigenvalue of an $n \times n$ matrix A, then the dimension of the eigenspace corresponding to λ_0 is called the ***geometric multiplicity*** of λ_0, and the number of times that $\lambda - \lambda_0$ appears as a factor in the characteristic polynomial of A is called the ***algebraic multiplicity*** of λ_0. The following theorem, which we state without proof, summarizes the preceding discussion.

Theorem 5.2.4

Geometric and Algebraic Multiplicity

If A is a square matrix, then:

(a) For every eigenvalue of A, the geometric multiplicity is less than or equal to the algebraic multiplicity.

(b) A is diagonalizable if and only if its characteristic polynomial can be expressed as a product of linear factors, and the geometric multiplicity of every eigenvalue is equal to the algebraic multiplicity.

We will complete this section with an optional proof of Theorem 5.2.2(a).

OPTIONAL: Proof of Theorem 5.2.2(a) Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be eigenvectors of A corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. We will assume that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly dependent and obtain a contradiction. We can then conclude that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly independent.

Since an eigenvector is nonzero by definition, $\{\mathbf{v}_1\}$ is linearly independent. Let r be the largest integer such that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is linearly independent. Since we are assuming that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is linearly dependent, r satisfies $1 \leq r < k$. Moreover, by the definition of r, the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{r+1}\}$ is linearly dependent. Thus, there are scalars $c_1, c_2, \ldots, c_{r+1}$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_{r+1}\mathbf{v}_{r+1} = \mathbf{0} \tag{5}$$

Multiplying both sides of (5) by A and using the fact that

$$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1, \quad A\mathbf{v}_2 = \lambda_2\mathbf{v}_2, \ldots, \quad A\mathbf{v}_{r+1} = \lambda_{r+1}\mathbf{v}_{r+1}$$

we obtain

$$c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_{r+1}\lambda_{r+1}\mathbf{v}_{r+1} = \mathbf{0} \tag{6}$$

If we now multiply both sides of (5) by λ_{r+1} and subtract the resulting equation from (6) we obtain

$$c_1(\lambda_1 - \lambda_{r+1})\mathbf{v}_1 + c_2(\lambda_2 - \lambda_{r+1})\mathbf{v}_2 + \cdots + c_r(\lambda_r - \lambda_{r+1})\mathbf{v}_r = \mathbf{0}$$

Since $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a linearly independent set, this equation implies that

$$c_1(\lambda_1 - \lambda_{r+1}) = c_2(\lambda_2 - \lambda_{r+1}) = \cdots = c_r(\lambda_r - \lambda_{r+1}) = 0$$

and since $\lambda_1, \lambda_2, \ldots, \lambda_{r+1}$ are assumed to be distinct, it follows that

$$c_1 = c_2 = \cdots = c_r = 0 \tag{7}$$

Substituting these values in (5) yields

$$c_{r+1}\mathbf{v}_{r+1} = \mathbf{0}$$

Since the eigenvector $\mathbf{v}_{r+1}$ is nonzero, it follows that

$$c_{r+1} = 0 \tag{8}$$

But equations (7) and (8) contradict the fact that $c_1, c_2, \ldots, c_{r+1}$ are not all zero, so the proof is complete.

Exercise Set 5.2

*In Exercises **1–4**, show that A and B are not similar matrices.*

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

2. $A = \begin{bmatrix} 4 & -1 \\ 2 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 4 & 1 \\ 2 & 4 \end{bmatrix}$

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

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

*In Exercises **5–8**, find a matrix P that diagonalizes A, and check your work by computing $P^{-1}AP$.*

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

6. $A = \begin{bmatrix} -14 & 12 \\ -20 & 17 \end{bmatrix}$

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

8. $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$

9. Let

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

a. Find the eigenvalues of A.

b. For each eigenvalue λ, find the rank of the matrix $\lambda I - A$.

c. Is A diagonalizable? Justify your conclusion.

10. Follow the directions in Exercise 9 for the matrix

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

*In Exercises **11–14**, find the geometric and algebraic multiplicity of each eigenvalue of the matrix A, and determine whether A is diagonalizable. If A is diagonalizable, then find a matrix P that diagonalizes A, and find $P^{-1}AP$.*

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

12. $A = \begin{bmatrix} 19 & -9 & -6 \\ 25 & -11 & -9 \\ 17 & -9 & -4 \end{bmatrix}$

13. $A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 3 & 0 & 1 \end{bmatrix}$

14. $A = \begin{bmatrix} 5 & 0 & 0 \\ 1 & 5 & 0 \\ 0 & 1 & 5 \end{bmatrix}$

*In each part of Exercises **15–16**, the characteristic equation of a matrix A is given. Find the size of the matrix and the possible dimensions of its eigenspaces.*

15. a. $(\lambda - 1)(\lambda + 3)(\lambda - 5) = 0$

b. $\lambda^2(\lambda - 1)(\lambda - 2)^3 = 0$

16. a. $\lambda^3(\lambda^2 - 5\lambda - 6) = 0$

b. $\lambda^3 - 3\lambda^2 + 3\lambda - 1 = 0$

*In Exercises **17–18**, use the method of Example 6 to compute the matrix A^{10}.*

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

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

19. Let

$$A = \begin{bmatrix} -1 & 7 & -1 \\ 0 & 1 & 0 \\ 0 & 15 & -2 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 5 \end{bmatrix}$$

Confirm that P diagonalizes A, and then compute A^{11}.

20. Let

$$A = \begin{bmatrix} 1 & -2 & 8 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 1 & -4 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Confirm that P diagonalizes A, and then compute each of the following powers of A.

a. A^{1000} **b.** A^{-1000} **c.** A^{2301} **d.** A^{-2301}

21. Find A^n if n is a positive integer and

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

22. Show that the matrices

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

are similar.

23. We know from Table 1 that similar matrices have the same rank. Show that the converse is false by showing that the matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

have the same rank but are not similar. [*Suggestion:* If they were similar, then there would be an invertible 2×2 matrix P for which $AP = PB$. Show that there is no such matrix.]

24. We know from Table 1 that similar matrices have the same eigenvalues. Use the method of Exercise 23 to show that the converse is false by showing that the matrices

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

have the same eigenvalues but are not similar.

25. If A, B, and C are $n \times n$ matrices such that A is similar to B and B is similar to C, do you think that A must be similar to C? Justify your answer.

26. a. Is it possible for an $n \times n$ matrix to be similar to itself? Justify your answer.

b. What can you say about an $n \times n$ matrix that is similar to $0_{n \times n}$? Justify your answer.

c. Is it possible for a nonsingular matrix to be similar to a singular matrix? Justify your answer.

27. Suppose that the characteristic polynomial of some matrix A is found to be $p(\lambda) = (\lambda - 1)(\lambda - 3)^2(\lambda - 4)^3$. In each part, answer the question and explain your reasoning.

a. What can you say about the dimensions of the eigenspaces of A?

b. What can you say about the dimensions of the eigenspaces if you know that A is diagonalizable?

c. If $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly independent set of eigenvectors of A, all of which correspond to the same eigenvalue of A, what can you say about that eigenvalue?

28. Let

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

Show that

a. A is diagonalizable if $(a - d)^2 + 4bc > 0$.

b. A is not diagonalizable if $(a - d)^2 + 4bc < 0$.

[*Hint:* See Exercise 29 of Section 5.1.]

29. In the case where the matrix A in Exercise 28 is diagonalizable, find a matrix P that diagonalizes A. [*Hint:* See Exercise 30 of Section 5.1.]

In Exercises 30–33, find the standard matrix A for the given linear operator, and determine whether that matrix is diagonalizable. If diagonalizable, find a matrix P that diagonalizes A.

30. $T(x_1, x_2) = (2x_1 - x_2, x_1 + x_2)$

31. $T(x_1, x_2) = (-x_2, -x_1)$

32. $T(x_1, x_2, x_3) = (8x_1 + 3x_2 - 4x_3, -3x_1 + x_2 + 3x_3, 4x_1 + 3x_2)$

33. $T(x_1, x_2, x_3) = (3x_1, x_2, x_1 - x_2)$

34. If P is a fixed $n \times n$ matrix, then the similarity transformation

$$A \to P^{-1}AP$$

can be viewed as an operator $S_P(A) = P^{-1}AP$ on the vector space M_{nn} of $n \times n$ matrices.

a. Show that S_P is a linear operator.

b. Find the kernel of S_P.

c. Find the rank of S_P.

Working with Proofs

35. Prove that similar matrices have the same rank and nullity.

36. Prove that similar matrices have the same trace.

37. Prove that if A is diagonalizable, then so is A^k for every positive integer k.

38. We know from Table 1 that similar matrices, A and B, have the same eigenvalues. However, it is not true that those eigenvalues have the same corresponding eigenvectors for the two matrices. Prove that if $B = P^{-1}AP$, and $\mathbf{v}$ is an eigenvector of B corresponding to the eigenvalue λ, then $P\mathbf{v}$ is the eigenvector of A corresponding to λ.

39. Let A be an $n \times n$ matrix, and let $q(A)$ be the matrix

$$q(A) = a_n A^n + a_{n-1} A^{n-1} + \cdots + a_1 A + a_0 I_n$$

a. Prove that if $B = P^{-1}AP$, then $q(B) = P^{-1}q(A)P$.

b. Prove that if A is diagonalizable, then so is $q(A)$.

40. Prove that if A is a diagonalizable matrix, then the rank of A is the number of nonzero eigenvalues of A.

41. This problem will lead you through a proof of the fact that the algebraic multiplicity of an eigenvalue of an $n \times n$ matrix A is greater than or equal to the geometric multiplicity. For this purpose, assume that λ_0 is an eigenvalue with geometric multiplicity k.

a. Prove that there is a basis $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ for R^n in which the first k vectors of B form a basis for the eigenspace corresponding to λ_0.

b. Let P be the matrix having the vectors in B as columns. Prove that the product AP can be expressed as

$$AP = P\begin{bmatrix} \lambda_0 I_k & X \\ 0 & Y \end{bmatrix}$$

[*Hint:* Compare the first k column vectors on both sides.]

c. Use the result in part (b) to prove that A is similar to

$$C = \begin{bmatrix} \lambda_0 I_k & X \\ 0 & Y \end{bmatrix}$$

and hence that A and C have the same characteristic polynomial.

d. By considering $\det(\lambda I - C)$, prove that the characteristic polynomial of C (and hence A) contains the factor $(\lambda - \lambda_0)$ at least k times, thereby proving that the algebraic multiplicity of λ_0 is greater than or equal to the geometric multiplicity k.

True-False Exercises

TF. In parts (a)–(i) determine whether the statement is true or false, and justify your answer.

a. An $n \times n$ matrix with fewer than n distinct eigenvalues is not diagonalizable.

b. An $n \times n$ matrix with fewer than n linearly independent eigenvectors is not diagonalizable.

c. If A and B are similar $n \times n$ matrices, then there exists an invertible $n \times n$ matrix P such that $PA = BP$.

d. If A is diagonalizable, then there is a unique matrix P such that $P^{-1}AP$ is diagonal.

e. If A is diagonalizable and invertible, then A^{-1} is diagonalizable.

f. If A is diagonalizable, then A^T is diagonalizable.

g. If there is a basis for R^n consisting of eigenvectors of an $n \times n$ matrix A, then A is diagonalizable.

h. If every eigenvalue of a matrix A has algebraic multiplicity 1, then A is diagonalizable.

i. If 0 is an eigenvalue of a matrix A, then A^2 is singular.

Working with Technology

T1. Generate a random 4×4 matrix A and an invertible 4×4 matrix P and then confirm, as stated in Table 1, that $P^{-1}AP$ and A have the same

a. determinant.

b. rank.

c. nullity.

d. trace.

e. characteristic polynomial.

f. eigenvalues.

T2. **a.** Use Theorem 5.2.1 to show that the following matrix is diagonalizable.

$$A = \begin{bmatrix} -13 & -60 & -60 \\ 10 & 42 & 40 \\ -5 & -20 & -18 \end{bmatrix}$$

b. Find a matrix P that diagonalizes A.

c. Use the method of Example 6 to compute A^{10}, and check your result by computing A^{10} directly.

T3. Use Theorem 5.2.1 to show that the following matrix is not diagonalizable.

$$A = \begin{bmatrix} -10 & 11 & -6 \\ -15 & 16 & -10 \\ -3 & 3 & -2 \end{bmatrix}$$

5.3 Complex Vector Spaces

Because the characteristic equation of a square matrix can have complex solutions, the notions of complex eigenvalues and eigenvectors arise naturally, even within the context of matrices with real entries. In this section we will discuss this idea and use our results to study symmetric matrices in more detail. A review of the essentials of complex numbers appears in the back of this text.

Review of Complex Numbers

Recall that if $z = a + bi$ is a complex number, then:

- $\text{Re}(z) = a$ and $\text{Im}(z) = b$ are called the ***real part*** of z and the ***imaginary part*** of z, respectively,
- $|z| = \sqrt{a^2 + b^2}$ is called the ***modulus*** (or ***absolute value***) of z,
- $\bar{z} = a - bi$ is called the ***complex conjugate*** of z,
- $z\bar{z} = a^2 + b^2 = |z|^2$,
- the angle ϕ in Figure 5.3.1 is called an ***argument*** of z,
- $\text{Re}(z) = |z|\cos\phi$,
- $\text{Im}(z) = |z|\sin\phi$,
- $z = |z|(\cos\phi + i\sin\phi)$ is called the ***polar form*** of z.

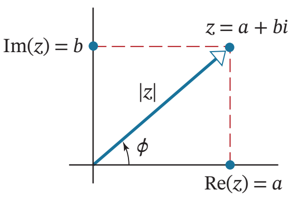


FIGURE 5.3.1

Complex Eigenvalues

In Formula (3) of Section 5.1 we observed that the characteristic equation of a general $n \times n$ matrix A has the form

$$\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0 \tag{1}$$

in which the highest power of λ has a coefficient of 1. Up to now we have limited our discussion to matrices in which the solutions of (1) are real numbers. However, it is possible for the characteristic equation of a matrix A with real entries to have imaginary solutions; for example, the characteristic equation of the matrix

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

is

$$\begin{vmatrix} \lambda + 2 & 1 \\ -5 & \lambda - 2 \end{vmatrix} = \lambda^2 + 1 = 0$$

which has the imaginary solutions $\lambda = i$ and $\lambda = -i$. To deal with this case we will need to explore the notion of a complex vector space and some related ideas.

Vectors in C^n

A vector space in which scalars are allowed to be complex numbers is called a ***complex vector space***. In this section we will be concerned only with the following complex generalization of the real vector space R^n.

Definition 1

If n is a positive integer, then a ***complex n-tuple*** is a sequence of n complex numbers $(v_1, v_2, \ldots, v_n)$. The set of all complex n-tuples is called ***complex n-space*** and is denoted by C^n. Scalars are complex numbers, and the operations of addition, subtraction, and scalar multiplication are performed componentwise.

The terminology used for n-tuples of real numbers applies to complex n-tuples without change. Thus, if $v_1, v_2, \ldots, v_n$ are complex numbers, then we call $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ a ***vector*** in C^n and $v_1, v_2, \ldots, v_n$ its ***components***. Some examples of vectors in C^3 are

$$\mathbf{u} = (1 + i, -4i, 3 + 2i), \quad \mathbf{v} = (0, i, 5), \quad \mathbf{w} = \left(6 - \sqrt{2}\,i, 9 + \tfrac{1}{2}i, \pi i\right)$$

Every vector

$$\mathbf{v} = (v_1, v_2, \ldots, v_n) = (a_1 + b_1 i, a_2 + b_2 i, \ldots, a_n + b_n i)$$

in C^n can be split into ***real*** and ***imaginary parts*** as

$$\mathbf{v} = (a_1, a_2, \ldots, a_n) + i\,(b_1, b_2, \ldots, b_n)$$

which we also denote as

$$\mathbf{v} = \operatorname{Re}(\mathbf{v}) + i\operatorname{Im}(\mathbf{v})$$

where

$$\operatorname{Re}(\mathbf{v}) = (a_1, a_2, \ldots, a_n) \quad \text{and} \quad \operatorname{Im}(\mathbf{v}) = (b_1, b_2, \ldots, b_n)$$

The vector

$$\overline{\mathbf{v}} = (\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_n) = (a_1 - b_1 i, a_2 - b_2 i, \ldots, a_n - b_n i)$$

is called the ***complex conjugate*** of $\mathbf{v}$ and can be expressed in terms of $\operatorname{Re}(\mathbf{v})$ and $\operatorname{Im}(\mathbf{v})$ as

$$\boxed{\overline{\mathbf{v}} = (a_1, a_2, \ldots, a_n) - i\,(b_1, b_2, \ldots, b_n) = \operatorname{Re}(\mathbf{v}) - i\operatorname{Im}(\mathbf{v})} \tag{2}$$

It follows that the vectors in R^n can be viewed as those vectors in C^n whose imaginary part is zero; or stated another way, a vector $\mathbf{v}$ in C^n is in R^n if and only if $\overline{\mathbf{v}} = \mathbf{v}$.

In this section we will need to distinguish between matrices whose entries *must* be real numbers, called ***real matrices***, and matrices whose entries may be *either* real numbers or complex numbers, called ***complex matrices***. When convenient, you can think of a real matrix as a complex matrix each of whose entries has a zero imaginary part. The standard operations on real matrices carry over without change to complex matrices, and all of the familiar properties of matrices continue to hold.

If A is a complex matrix, then $\operatorname{Re}(A)$ and $\operatorname{Im}(A)$ are the matrices formed from the real and imaginary parts of the entries of A, and $\overline{A}$ is the matrix formed by taking the complex conjugate of each entry in A.

EXAMPLE 1 | Real and Imaginary Parts of Vectors and Matrices

Let

$$\mathbf{v} = (3+i, -2i, 5) \quad \text{and} \quad A = \begin{bmatrix} 1+i & -i \\ 4 & 6-2i \end{bmatrix}$$

Then

$$\overline{\mathbf{v}} = (3-i, 2i, 5), \quad \operatorname{Re}(\mathbf{v}) = (3, 0, 5), \quad \operatorname{Im}(\mathbf{v}) = (1, -2, 0)$$

$$\overline{A} = \begin{bmatrix} 1-i & i \\ 4 & 6+2i \end{bmatrix}, \quad \operatorname{Re}(A) = \begin{bmatrix} 1 & 0 \\ 4 & 6 \end{bmatrix}, \quad \operatorname{Im}(A) = \begin{bmatrix} 1 & -1 \\ 0 & -2 \end{bmatrix}$$

$$\det(A) = \begin{vmatrix} 1+i & -i \\ 4 & 6-2i \end{vmatrix} = (1+i)(6-2i) - (-i)(4) = 8 + 8i$$

As you might expect, if A is a complex matrix, then A and $\overline{A}$ can be expressed in terms of $\operatorname{Re}(A)$ and $\operatorname{Im}(A)$ as

$$A = \operatorname{Re}(A) + i\operatorname{Im}(A)$$
$$\overline{A} = \operatorname{Re}(A) - i\operatorname{Im}(A)$$

Algebraic Properties of the Complex Conjugate

The next two theorems list some properties of complex vectors and matrices that we will need in this section. Some of the proofs are given as exercises.

Theorem 5.3.1

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in C^n, and if k is a scalar, then:

(a) $\overline{\overline{\mathbf{u}}} = \mathbf{u}$

(b) $\overline{k\mathbf{u}} = \overline{k}\,\overline{\mathbf{u}}$

(c) $\overline{\mathbf{u}+\mathbf{v}} = \overline{\mathbf{u}} + \overline{\mathbf{v}}$

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

Theorem 5.3.2

If A is an $m \times k$ complex matrix and B is a $k \times n$ complex matrix, then:

(a) $\overline{\overline{A}} = A$

(b) $\overline{(A^T)} = (\overline{A})^T$

(c) $\overline{AB} = \overline{A}\,\overline{B}$

The Complex Euclidean Inner Product

The following definition extends the notions of dot product and norm to C^n.

Definition 2

If $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ are vectors in C^n, then the ***complex Euclidean inner product*** of $\mathbf{u}$ and $\mathbf{v}$ (also called the ***complex dot product***) is denoted by $\mathbf{u} \cdot \mathbf{v}$ and is defined as

$$\boxed{\mathbf{u} \cdot \mathbf{v} = u_1\overline{v}_1 + u_2\overline{v}_2 + \cdots + u_n\overline{v}_n} \tag{3}$$

We also define the ***Euclidean norm*** on C^n to be

$$\boxed{\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2}} \tag{4}$$

The complex conjugates in (3) ensure that $\|\mathbf{v}\|$ is a real number, for without them the quantity $\mathbf{v} \cdot \mathbf{v}$ in (4) might be imaginary.

As in the real case, we call $\mathbf{v}$ a ***unit vector*** in C^n if $\|\mathbf{v}\| = 1$, and we say two vectors $\mathbf{u}$ and $\mathbf{v}$ are ***orthogonal*** if $\mathbf{u} \cdot \mathbf{v} = 0$.

EXAMPLE 2 | Complex Euclidean Inner Product and Norm

Find $\mathbf{u} \cdot \mathbf{v}$, $\mathbf{v} \cdot \mathbf{u}$, $\|\mathbf{u}\|$, and $\|\mathbf{v}\|$ for the vectors

$$\mathbf{u} = (1+i, i, 3-i) \quad \text{and} \quad \mathbf{v} = (1+i, 2, 4i)$$

Solution

$$\mathbf{u} \cdot \mathbf{v} = (1+i)(\overline{1+i}) + i(\overline{2}) + (3-i)(\overline{4i}) = (1+i)(1-i) + 2i + (3-i)(-4i) = -2 - 10i$$

$$\mathbf{v} \cdot \mathbf{u} = (1+i)(\overline{1+i}) + 2(\overline{i}) + (4i)(\overline{3-i}) = (1+i)(1-i) - 2i + 4i(3+i) = -2 + 10i$$

$$\|\mathbf{u}\| = \sqrt{|1+i|^2 + |i|^2 + |3-i|^2} = \sqrt{2+1+10} = \sqrt{13}$$

$$\|\mathbf{v}\| = \sqrt{|1+i|^2 + |2|^2 + |4i|^2} = \sqrt{2+4+16} = \sqrt{22}$$

Example 2 reveals a major difference between the dot product on R^n and the complex dot product on C^n. For the dot product on R^n we always have $\mathbf{v} \cdot \mathbf{u} = \mathbf{u} \cdot \mathbf{v}$ (the *symmetry property*), but for the complex dot product the corresponding relationship is given by $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$, which is called its ***antisymmetry*** property. The following theorem is an analog of Theorem 3.2.2.

Theorem 5.3.3

If $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ are vectors in C^n, and if k is a scalar, then the complex Euclidean inner product has the following properties:

(*a*) $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$ **[Antisymmetry property]**

(*b*) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$ **[Distributive property]**

(*c*) $k(\mathbf{u} \cdot \mathbf{v}) = (k\mathbf{u}) \cdot \mathbf{v}$ **[Homogeneity property]**

(*d*) $\mathbf{u} \cdot k\mathbf{v} = \overline{k}(\mathbf{u} \cdot \mathbf{v})$ **[Antihomogeneity property]**

(*e*) $\mathbf{v} \cdot \mathbf{v} \geq 0$ and $\mathbf{v} \cdot \mathbf{v} = 0$ if and only if $\mathbf{v} = \mathbf{0}$ **[Positivity property]**

Parts (*c*) and (*d*) of this theorem state that a scalar multiplying a complex Euclidean inner product can be regrouped with the first vector, but to regroup it with the second vector you must first take its complex conjugate. We will prove part (*d*), and leave the others as exercises.

***Proof* (*d*)**

$$k(\mathbf{u}\cdot\mathbf{v}) = k(\overline{\mathbf{v}\cdot\mathbf{u}}) = \overline{\overline{k}}\,(\overline{\mathbf{v}\cdot\mathbf{u}}) = \overline{\overline{k}\,(\mathbf{v}\cdot\mathbf{u})} = \overline{(\overline{k}\mathbf{v})\cdot\mathbf{u}} = \mathbf{u}\cdot(\overline{k}\mathbf{v})$$

To complete the proof, substitute $\overline{k}$ for k and use the fact that $\overline{\overline{k}} = k$. ■

Recall from Table 1 of Section 3.2 that if $\mathbf{u}$ and $\mathbf{v}$ are *column vectors* in R^n, then their dot product can be expressed as

$$\mathbf{u}\cdot\mathbf{v} = \mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u}$$

The analogous formulas in C^n are (verify)

$$\boxed{\mathbf{u}\cdot\mathbf{v} = \mathbf{u}^T\overline{\mathbf{v}} = \overline{\mathbf{v}}^T\mathbf{u}} \tag{5}$$

Vector Concepts in C^n

Except for the use of complex scalars, the notions of linear combination, linear independence, subspace, spanning, basis, and dimension carry over without change to C^n.

Is R^n a subspace of C^n? Explain.

Eigenvalues and eigenvectors are defined for complex matrices exactly as for real matrices: If A is an $n \times n$ matrix with complex entries, then the complex roots of the characteristic equation $\det(\lambda I - A) = 0$ are called ***complex eigenvalues*** of A. As in the real case, λ is a complex eigenvalue of A if and only if there exists a nonzero vector $\mathbf{x}$ in C^n such that $A\mathbf{x} = \lambda\mathbf{x}$. Each such $\mathbf{x}$ is called a ***complex eigenvector*** of A corresponding to λ. The complex eigenvectors of A corresponding to λ are the nonzero solutions of the linear system $(\lambda I - A)\mathbf{x} = \mathbf{0}$, and the set of all such solutions is a subspace of C^n, called the ***complex eigenspace*** of A corresponding to λ.

The following theorem states that if a *real matrix* has complex eigenvalues, then those eigenvalues and their corresponding eigenvectors occur in conjugate pairs.

Theorem 5.3.4

If λ is an eigenvalue of a real $n \times n$ matrix A, and if $\mathbf{x}$ is a corresponding eigenvector, then $\overline{\lambda}$ is also an eigenvalue of A, and $\overline{\mathbf{x}}$ is a corresponding eigenvector.

Proof Since λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector, we have

$$\overline{A\mathbf{x}} = \overline{\lambda\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}} \tag{6}$$

However, $\overline{A} = A$, since A has real entries, so it follows from part (*c*) of Theorem 5.3.2 that

$$\overline{A\mathbf{x}} = \overline{A}\overline{\mathbf{x}} = A\overline{\mathbf{x}} \tag{7}$$

Equations (6) and (7) together imply that

$$A\overline{\mathbf{x}} = \overline{A\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$$

in which $\overline{\mathbf{x}} \neq \mathbf{0}$ (why?). This tells us that $\overline{\lambda}$ is an eigenvalue of A and $\overline{\mathbf{x}}$ is a corresponding eigenvector. ■

EXAMPLE 3 | Complex Eigenvalues and Eigenvectors

Find the eigenvalues and bases for the eigenspaces of

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

Solution The characteristic polynomial of A is

$$\begin{vmatrix} \lambda + 2 & 1 \\ -5 & \lambda - 2 \end{vmatrix} = \lambda^2 + 1 = (\lambda - i)(\lambda + i)$$

so the eigenvalues of A are $\lambda = i$ and $\lambda = -i$. Note that these eigenvalues are complex conjugates, as guaranteed by Theorem 5.3.4. To find the eigenvectors we must solve the system

$$\begin{bmatrix} \lambda + 2 & 1 \\ -5 & \lambda - 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

with $\lambda = i$ and then with $\lambda = -i$. With $\lambda = i$, this system becomes

$$\begin{bmatrix} i + 2 & 1 \\ -5 & i - 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{8}$$

We could solve this system by reducing the augmented matrix

$$\begin{bmatrix} i + 2 & 1 & 0 \\ -5 & i - 2 & 0 \end{bmatrix} \tag{9}$$

to reduced row echelon form by Gauss–Jordan elimination, though the complex arithmetic is somewhat tedious. A simpler procedure here is first to observe that the reduced row echelon form of (9) must have a row of zeros because (8) has nontrivial solutions. This being the case, each row of (9) must be a scalar multiple of the other, and hence the first row can be made into a row of zeros by adding a suitable multiple of the second row to it. Accordingly, we can simply set the entries in the first row to zero, then interchange the rows, and then multiply the new first row by $-\frac{1}{5}$ to obtain the reduced row echelon form

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

Thus, a general solution of the system is

$$x_1 = \left(-\tfrac{2}{5} + \tfrac{1}{5}i\right)t, \quad x_2 = t$$

This tells us that the eigenspace corresponding to $\lambda = i$ is one-dimensional and consists of all complex scalar multiples of the basis vector

$$\mathbf{x} = \begin{bmatrix} -\frac{2}{5} + \frac{1}{5}i \\ 1 \end{bmatrix} \tag{10}$$

As a check, let us confirm that $A\mathbf{x} = i\mathbf{x}$. We obtain

$$\begin{aligned} A\mathbf{x} &= \begin{bmatrix} -2 & -1 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} -\frac{2}{5} + \frac{1}{5}i \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -2\left(-\frac{2}{5} + \frac{1}{5}i\right) - 1 \\ 5\left(-\frac{2}{5} + \frac{1}{5}i\right) + 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{5} - \frac{2}{5}i \\ i \end{bmatrix} = i\mathbf{x} \end{aligned}$$

We could find a basis for the eigenspace corresponding to $\lambda = -i$ in a similar way, but the work is unnecessary since Theorem 5.3.4 implies that

$$\overline{\mathbf{x}} = \begin{bmatrix} -\frac{2}{5} - \frac{1}{5}i \\ 1 \end{bmatrix} \tag{11}$$

must be a basis for this eigenspace. The following computations confirm that $\overline{\mathbf{x}}$ is an eigenvector of A corresponding to $\lambda = -i$:

$$\begin{aligned} A\overline{\mathbf{x}} &= \begin{bmatrix} -2 & -1 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} -\frac{2}{5} - \frac{1}{5}i \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -2\left(-\frac{2}{5} - \frac{1}{5}i\right) - 1 \\ 5\left(-\frac{2}{5} - \frac{1}{5}i\right) + 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{5} + \frac{2}{5}i \\ -i \end{bmatrix} = -i\overline{\mathbf{x}} \end{aligned}$$

Since a number of our subsequent examples will involve 2×2 matrices with real entries, it will be useful to discuss some general results about the eigenvalues of such matrices. Observe first that the characteristic polynomial of the matrix

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

is

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - a & -b \\ -c & \lambda - d \end{vmatrix} = (\lambda - a)(\lambda - d) - bc = \lambda^2 - (a+d)\lambda + (ad - bc)$$

We can express this in terms of the trace and determinant of A as

$$\det(\lambda I - A) = \lambda^2 - \operatorname{tr}(A)\lambda + \det(A) \tag{12}$$

from which it follows that the characteristic equation of A is

$$\boxed{\lambda^2 - \operatorname{tr}(A)\lambda + \det(A) = 0} \tag{13}$$

Now recall from algebra that if $ax^2 + bx + c = 0$ is a quadratic equation with real coefficients, then the ***discriminant*** $b^2 - 4ac$ determines the nature of the roots:

$$\begin{aligned} b^2 - 4ac &> 0 \quad \textbf{[Two distinct real roots]} \\ b^2 - 4ac &= 0 \quad \textbf{[One repeated real root]} \\ b^2 - 4ac &< 0 \quad \textbf{[Two conjugate imaginary roots]} \end{aligned}$$

Applying this to (13) with $a = 1, b = -\operatorname{tr}(A)$, and $c = \det(A)$ yields the following theorem.

Theorem 5.3.5

If A is a 2×2 matrix with real entries, then the characteristic equation of A is $\lambda^2 - \operatorname{tr}(A)\lambda + \det(A) = 0$ and

(*a*) A has two distinct real eigenvalues if $\operatorname{tr}(A)^2 - 4\det(A) > 0$;

(*b*) A has one repeated real eigenvalue if $\operatorname{tr}(A)^2 - 4\det(A) = 0$;

(*c*) A has two complex conjugate eigenvalues if $\operatorname{tr}(A)^2 - 4\det(A) < 0$.

EXAMPLE 4 | Eigenvalues of a 2×2 Matrix

In each part, use Formula (13) for the characteristic equation to find the eigenvalues of

$$(a)\ A = \begin{bmatrix} 2 & 2 \\ -1 & 5 \end{bmatrix} \quad (b)\ A = \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \quad (c)\ A = \begin{bmatrix} 2 & 3 \\ -3 & 2 \end{bmatrix}$$

Solution (*a*) We have $\operatorname{tr}(A) = 7$ and $\det(A) = 12$, so the characteristic equation of A is

$$\lambda^2 - 7\lambda + 12 = 0$$

Factoring yields $(\lambda - 4)(\lambda - 3) = 0$, so the eigenvalues of A are $\lambda = 4$ and $\lambda = 3$.

Solution (*b*) We have $\operatorname{tr}(A) = 2$ and $\det(A) = 1$, so the characteristic equation of A is

$$\lambda^2 - 2\lambda + 1 = 0$$

Factoring this equation yields $(\lambda - 1)^2 = 0$, so $\lambda = 1$ is the only eigenvalue of A; it has algebraic multiplicity 2.

Solution (*c*) We have $\operatorname{tr}(A) = 4$ and $\det(A) = 13$, so the characteristic equation of A is

$$\lambda^2 - 4\lambda + 13 = 0$$

Solving this equation by the quadratic formula yields

$$\lambda = \frac{4 \pm \sqrt{(-4)^2 - 4(13)}}{2} = \frac{4 \pm \sqrt{-36}}{2} = 2 \pm 3i$$

Thus, the eigenvalues of A are $\lambda = 2 + 3i$ and $\lambda = 2 - 3i$.

Historical Note

Olga Taussky-Todd (1906–1995)

Olga Taussky-Todd was one of the pioneering women in matrix analysis and the first woman appointed to the faculty at the California Institute of Technology. She worked at the National Physical Laboratory in London during World War II, where she was assigned to study flutter in supersonic aircraft. While there, she realized that some results about the eigenvalues of a certain 6×6 complex matrix could be used to answer key questions about the flutter problem that would otherwise have required laborious calculation. After World War II Olga Taussky-Todd continued her work on matrix-related subjects and helped to draw many known but disparate results about matrices into the coherent subject that we now call matrix theory.

[Image: Courtesy of the Archives, California Institute of Technology]

Symmetric Matrices Have Real Eigenvalues

Our next result, which is concerned with the eigenvalues of real symmetric matrices, is important in a wide variety of applications. The key to its proof is to think of a real symmetric matrix as a complex matrix whose entries have an imaginary part of zero.

Theorem 5.3.6

If A is a real symmetric matrix, then A has real eigenvalues.

Proof Suppose that λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector, where we allow for the possibility that λ is complex and $\mathbf{x}$ is in C^n. Thus,

$$A\mathbf{x} = \lambda\mathbf{x}$$

where $\mathbf{x} \neq \mathbf{0}$. If we multiply both sides of this equation by $\overline{\mathbf{x}}^T$ and use the fact that

$$\overline{\mathbf{x}}^T A\mathbf{x} = \overline{\mathbf{x}}^T(\lambda\mathbf{x}) = \lambda(\overline{\mathbf{x}}^T\mathbf{x}) = \lambda(\mathbf{x}\cdot\mathbf{x}) = \lambda\|\mathbf{x}\|^2$$

then we obtain

$$\lambda = \frac{\overline{\mathbf{x}}^T A\mathbf{x}}{\|\mathbf{x}\|^2}$$

Since the denominator in this expression is real, we can prove that λ is real by showing that

$$\overline{\overline{\mathbf{x}}^T A\mathbf{x}} = \overline{\mathbf{x}}^T A\mathbf{x} \tag{14}$$

But A is symmetric and has real entries, so it follows from the second equality in (5) and properties of the conjugate that

$$\overline{\overline{\mathbf{x}}^T A\mathbf{x}} = \overline{\overline{\mathbf{x}}}^T\,\overline{A\mathbf{x}} = \mathbf{x}^T\,\overline{A\mathbf{x}} = (\overline{A\mathbf{x}})^T\mathbf{x} = (\overline{A}\overline{\mathbf{x}})^T\mathbf{x} = (A\overline{\mathbf{x}})^T\mathbf{x} = \overline{\mathbf{x}}^T A^T\mathbf{x} = \overline{\mathbf{x}}^T A\mathbf{x} \quad \blacksquare$$

A Geometric Interpretation of Complex Eigenvalues

The following theorem is the key to understanding the geometric significance of complex eigenvalues of real 2×2 matrices.

Theorem 5.3.7

The eigenvalues of the real matrix

$$C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \tag{15}$$

are $\lambda = a \pm bi$. If a and b are not both zero, then this matrix can be factored as

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \tag{16}$$

where ϕ is the angle from the positive x-axis to the ray that joins the origin to the point (a, b) (**Figure 5.3.2**).

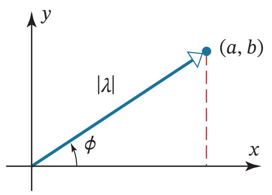


FIGURE 5.3.2

Geometrically, this theorem states that multiplication by a matrix of form (15) can be viewed as a rotation through the angle ϕ followed by a scaling with factor $|\lambda|$ (**Figure 5.3.3**).

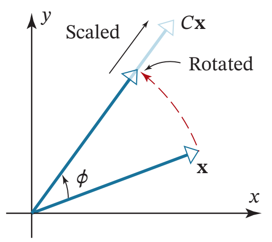


FIGURE 5.3.3

Proof The characteristic equation of C is $(\lambda - a)^2 + b^2 = 0$ (verify), from which it follows that the eigenvalues of C are $\lambda = a \pm bi$. Assuming that a and b are not both zero, let ϕ be the angle from the positive x-axis to the ray that joins the origin to the point (a, b). The angle ϕ is an argument of the eigenvalue $\lambda = a + bi$, so we see from Figure 5.3.2 that

$$a = |\lambda| \cos\phi \quad \text{and} \quad b = |\lambda| \sin\phi$$

It follows from this that the matrix in (15) can be written as

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \dfrac{a}{|\lambda|} & -\dfrac{b}{|\lambda|} \\ \dfrac{b}{|\lambda|} & \dfrac{a}{|\lambda|} \end{bmatrix} = \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \qquad \blacksquare$$

The following theorem, whose proof is considered in the exercises, shows that every real 2×2 matrix with complex eigenvalues is similar to a matrix of form (15).

Theorem 5.3.8

Let A be a real 2×2 matrix with complex eigenvalues $\lambda = a \pm bi$ (where $b \neq 0$). If $\mathbf{x}$ is an eigenvector of A corresponding to $\lambda = a - bi$, then the matrix $P = [\mathrm{Re}(\mathbf{x}) \;\; \mathrm{Im}(\mathbf{x})]$ is invertible and

$$A = P \begin{bmatrix} a & -b \\ b & a \end{bmatrix} P^{-1} \tag{17}$$

EXAMPLE 5 | A Matrix Factorization Using Complex Eigenvalues

Factor the matrix in Example 3 into form (17) using the eigenvalue $\lambda = -i$ and the corresponding eigenvector that was given in (11).

Solution For consistency with the notation in Theorem 5.3.8, let us denote the eigenvector in (11) that corresponds to $\lambda = -i$ by $\mathbf{x}$ (rather than $\bar{\mathbf{x}}$ as before). For this λ and $\mathbf{x}$ we have

$$a = 0, \quad b = 1, \quad \mathrm{Re}(\mathbf{x}) = \begin{bmatrix} -\frac{2}{5} \\ 1 \end{bmatrix}, \quad \mathrm{Im}(\mathbf{x}) = \begin{bmatrix} -\frac{1}{5} \\ 0 \end{bmatrix}$$

Thus,

$$P = [\mathrm{Re}(\mathbf{x}) \quad \mathrm{Im}(\mathbf{x})] = \begin{bmatrix} -\frac{2}{5} & -\frac{1}{5} \\ 1 & 0 \end{bmatrix}$$

so A can be factored in form (17) as

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

You may want to confirm this by multiplying out the right side.

A Geometric Interpretation of Theorem 5.3.8

To interpret what Theorem 5.3.8 says geometrically, let us denote the matrices on the right side of (16) by S and R_ϕ, respectively, and then use (16) to rewrite (17) as

$$A = PSR_\phi P^{-1} = P \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} P^{-1} \tag{18}$$

If we now view P as the transition matrix from the basis $B = \{\mathrm{Re}(\mathbf{x}), \mathrm{Im}(\mathbf{x})\}$ to the standard basis, then (18) tells us that computing a product $A\mathbf{x}_0$ can be broken down into a three-step process:

Interpreting Formula (18)

Step 1. Map $\mathbf{x}_0$ from standard coordinates into B-coordinates by forming the product $P^{-1}\mathbf{x}_0$.

Step 2. Rotate and scale the vector $P^{-1}\mathbf{x}_0$ by forming the product $SR_\phi P^{-1}\mathbf{x}_0$.

Step 3. Map the rotated and scaled vector back to standard coordinates to obtain $A\mathbf{x}_0 = PSR_\phi P^{-1}\mathbf{x}_0$.

Power Sequences

There are many problems in which one is interested in how successive applications of a matrix transformation affect a specific vector. For example, if A is the standard matrix for an operator on R^n and $\mathbf{x}_0$ is some fixed vector in R^n, then one might be interested in the behavior of the power sequence

$$\mathbf{x}_0, \quad A\mathbf{x}_0, \quad A^2\mathbf{x}_0, \ldots, \quad A^k\mathbf{x}_0, \ldots$$

For example, if

$$A = \begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ -\frac{3}{5} & \frac{11}{10} \end{bmatrix} \quad \text{and} \quad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

then with the help of a computer or calculator one can show that the first four terms in the power sequence are

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad A\mathbf{x}_0 = \begin{bmatrix} 1.25 \\ 0.5 \end{bmatrix}, \quad A^2\mathbf{x}_0 = \begin{bmatrix} 1.0 \\ -0.2 \end{bmatrix}, \quad A^3\mathbf{x}_0 = \begin{bmatrix} 0.35 \\ -0.82 \end{bmatrix}$$

With the help of MATLAB or a computer algebra system one can show that if the first 100 terms are plotted as ordered pairs (x, y), then the points move along the elliptical path shown in **Figure 5.3.4*a***.

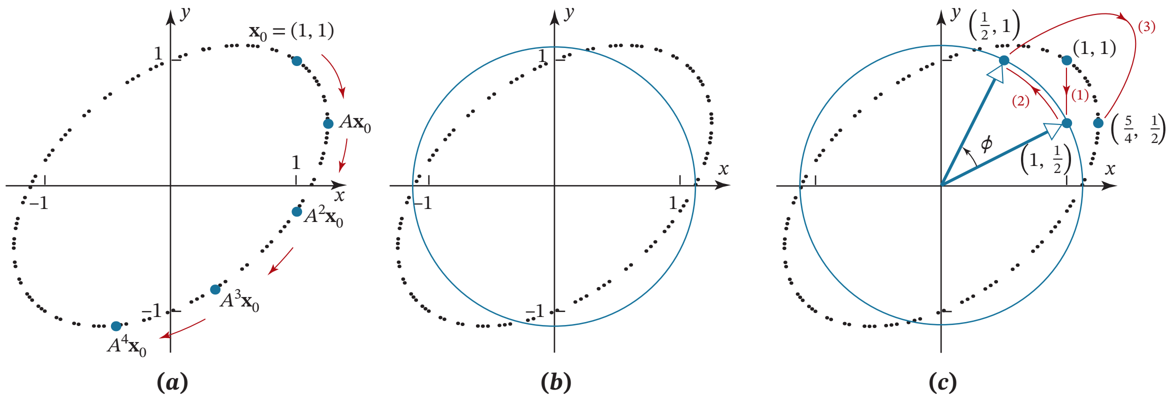


FIGURE 5.3.4

To understand why the points move along an elliptical path, we will need to examine the eigenvalues and eigenvectors of A. We leave it for you to show that the eigenvalues of A are $\lambda = \frac{4}{5} \pm \frac{3}{5}i$ and that the corresponding eigenvectors are

$$\lambda_1 = \tfrac{4}{5} - \tfrac{3}{5}i: \quad \mathbf{v}_1 = \left(\tfrac{1}{2} + i, 1\right) \quad \text{and} \quad \lambda_2 = \tfrac{4}{5} + \tfrac{3}{5}i: \quad \mathbf{v}_2 = \left(\tfrac{1}{2} - i, 1\right)$$

If we take $\lambda = \lambda_1 = \frac{4}{5} - \frac{3}{5}i$ and $\mathbf{x} = \mathbf{v}_1 = \left(\frac{1}{2} + i, 1\right)$ in (17) and use the fact that $|\lambda| = 1$, then we obtain the factorization

$$\underset{A}{\begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ -\frac{3}{5} & \frac{11}{10} \end{bmatrix}} = \underset{P}{\begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 0 \end{bmatrix}} \underset{R_\phi}{\begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix}} \underset{P^{-1}}{\begin{bmatrix} 0 & 1 \\ 1 & -\frac{1}{2} \end{bmatrix}} \tag{19}$$

where R_ϕ is a rotation about the origin through the angle ϕ whose tangent is

$$\tan\phi = \frac{\sin\phi}{\cos\phi} = \frac{3/5}{4/5} = \frac{3}{4} \qquad \left(\phi = \tan^{-1}\tfrac{3}{4} \approx 36.9^\circ\right)$$

The matrix P in (19) is the transition matrix from the basis

$$B = \{\text{Re}(\mathbf{x}), \text{Im}(\mathbf{x})\} = \left\{\left(\tfrac{1}{2}, 1\right), (1, 0)\right\}$$

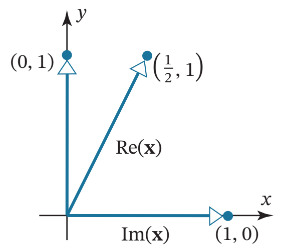


FIGURE 5.3.5

to the standard basis, and P^{-1} is the transition matrix from the standard basis to the basis B (**Figure 5.3.5**). Next, observe that if n is a positive integer, then (19) implies that

$$A^n\mathbf{x}_0 = (PR_\phi P^{-1})^n\mathbf{x}_0 = PR_\phi^n P^{-1}\mathbf{x}_0$$

so the product $A^n\mathbf{x}_0$ can be computed by first mapping $\mathbf{x}_0$ into the point $P^{-1}\mathbf{x}_0$ in B-coordinates, then multiplying by R_ϕ^n to rotate this point about the origin through the angle $n\phi$, and then multiplying $R_\phi^n P^{-1}\mathbf{x}_0$ by P to map the resulting point back to standard coordinates. We can now see what is happening geometrically: In B-coordinates each successive multiplication by A causes the point $P^{-1}\mathbf{x}_0$ to advance through an angle ϕ, thereby tracing a circular orbit about the origin. However, the basis B is *skewed* (not orthogonal), so when the points on the circular orbit are transformed back to standard coordinates, the effect is to distort the circular orbit into the elliptical orbit traced by $A^n\mathbf{x}_0$ (**Figure 5.3.4*b***).

Here are the computations for the first step (successive steps are illustrated in **Figure 5.3.4c**):

$$\begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ -\frac{3}{5} & \frac{11}{10} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix}\begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix} \qquad [\mathbf{x}_0 \text{ is mapped to } B\text{-coordinates.}]$$

$$= \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} \qquad [\text{The point } (1, \tfrac{1}{2}) \text{ is rotated through the angle } \phi.]$$

$$= \begin{bmatrix} \frac{5}{4} \\ \frac{1}{2} \end{bmatrix} \qquad [\text{The point } (\tfrac{1}{2}, 1) \text{ is mapped to standard coordinates.}]$$

Exercise Set 5.3

In Exercises 1–2, find $\overline{\mathbf{u}}$, $Re(\mathbf{u})$, $Im(\mathbf{u})$, *and* $\|\mathbf{u}\|$.

1. $\mathbf{u} = (2 - i, 4i, 1 + i)$

2. $\mathbf{u} = (6, 1 + 4i, 6 - 2i)$

In Exercises 3–4, show that $\mathbf{u}$, $\mathbf{v}$, *and* k *satisfy Theorem 5.3.1.*

3. $\mathbf{u} = (3 - 4i, 2 + i, -6i)$, $\mathbf{v} = (1 + i, 2 - i, 4)$, $k = i$

4. $\mathbf{u} = (6, 1 + 4i, 6 - 2i)$, $\mathbf{v} = (4, 3 + 2i, i - 3)$, $k = -i$

5. Solve the equation $i\mathbf{x} - 3\mathbf{v} = \overline{\mathbf{u}}$ for $\mathbf{x}$, where $\mathbf{u}$ and $\mathbf{v}$ are the vectors in Exercise 3.

6. Solve the equation $(1 + i)\mathbf{x} + 2\mathbf{u} = \overline{\mathbf{v}}$ for $\mathbf{x}$, where $\mathbf{u}$ and $\mathbf{v}$ are the vectors in Exercise 4.

In Exercises 7–8, find $\overline{A}$, $Re(A)$, $Im(A)$, $\det(A)$, *and* $tr(A)$.

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

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

9. Let A be the matrix given in Exercise 7, and let B be the matrix

$$B = \begin{bmatrix} 1 - i \\ 2i \end{bmatrix}$$

Confirm that these matrices have the properties stated in Theorem 5.3.2.

10. Let A be the matrix given in Exercise 8, and let B be the matrix

$$B = \begin{bmatrix} 5i \\ 1 - 4i \end{bmatrix}$$

Confirm that these matrices have the properties stated in Theorem 5.3.2.

In Exercises 11–12, compute $\mathbf{u}\cdot\mathbf{v}$, $\mathbf{u}\cdot\mathbf{w}$, *and* $\mathbf{v}\cdot\mathbf{w}$, *and show that the vectors satisfy Formula (5) and parts (a), (b), and (c) of Theorem 5.3.3.*

11. $\mathbf{u} = (i, 2i, 3)$, $\mathbf{v} = (4, -2i, 1 + i)$, $\mathbf{w} = (2 - i, 2i, 5 + 3i)$, $k = 2i$

12. $\mathbf{u} = (1 + i, 4, 3i)$, $\mathbf{v} = (3, -4i, 2 + 3i)$, $\mathbf{w} = (1 - i, 4i, 4 - 5i)$, $k = 1 + i$

13. Compute $\overline{(\mathbf{u}\cdot\overline{\mathbf{v}}) - \overline{\mathbf{w}\cdot\mathbf{u}}}$ for the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in Exercise 11.

14. Compute $\overline{(i\overline{\mathbf{u}}\cdot\mathbf{w}) + (\|\mathbf{u}\|\mathbf{v})}\cdot\mathbf{u}$ for the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in Exercise 12.

In Exercises 15–18, find the eigenvalues and bases for the eigenspaces of A.

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

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

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

18. $A = \begin{bmatrix} 8 & 6 \\ -3 & 2 \end{bmatrix}$

In Exercises 19–22, each matrix C *has form (15). Theorem 5.3.7 implies that* C *is the product of a scaling matrix with factor* $|\lambda|$ *and a rotation matrix with angle* ϕ. *Find* $|\lambda|$ *and* ϕ *for which* $-\pi < \phi \le \pi$.

19. $C = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

20. $C = \begin{bmatrix} 0 & 5 \\ -5 & 0 \end{bmatrix}$

21. $C = \begin{bmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 1 \end{bmatrix}$

22. $C = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ -\sqrt{2} & \sqrt{2} \end{bmatrix}$

In Exercises 23–26, find an invertible matrix P *and a matrix* C *of form (15) such that* $A = PCP^{-1}$.

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

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

25. $A = \begin{bmatrix} 8 & 6 \\ -3 & 2 \end{bmatrix}$

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

27. Find all complex scalars k, if any, for which $\mathbf{u}$ and $\mathbf{v}$ are orthogonal in C^3.

a. $\mathbf{u} = (2i, i, 3i)$, $\mathbf{v} = (i, 6i, k)$

b. $\mathbf{u} = (k, k, 1 + i)$, $\mathbf{v} = (1, -1, 1 - i)$

28. Show that if A is a real $n \times n$ matrix and $\mathbf{x}$ is a column vector in C^n, then $\mathrm{Re}(A\mathbf{x}) = A(\mathrm{Re}(\mathbf{x}))$ and $\mathrm{Im}(A\mathbf{x}) = A(\mathrm{Im}(\mathbf{x}))$.

29. The matrices

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

called ***Pauli spin matrices***, are used in quantum mechanics to study particle spin. The ***Dirac matrices***, which are also used in quantum mechanics, are expressed in terms of the Pauli spin matrices and the 2×2 identity matrix I_2 as

$$\beta = \begin{bmatrix} I_2 & 0 \\ 0 & -I_2 \end{bmatrix}, \quad \alpha_x = \begin{bmatrix} 0 & \sigma_1 \\ \sigma_1 & 0 \end{bmatrix},$$

$$\alpha_y = \begin{bmatrix} 0 & \sigma_2 \\ \sigma_2 & 0 \end{bmatrix}, \quad \alpha_z = \begin{bmatrix} 0 & \sigma_3 \\ \sigma_3 & 0 \end{bmatrix}$$

a. Show that $\beta^2 = \alpha_x^2 = \alpha_y^2 = \alpha_z^2$.

b. Matrices A and B for which $AB = -BA$ are said to be ***anticommutative***. Show that the Dirac matrices are anticommutative.

30. If k is a real scalar and $\mathbf{v}$ is a vector in R^n, then Theorem 3.2.1 states that $\|k\mathbf{v}\| = |k|\,\|\mathbf{v}\|$. Is this relationship also true if k is a complex scalar and $\mathbf{v}$ is a vector in C^n? Justify your answer.

Working with Proofs

31. Prove part (c) of Theorem 5.3.1.

32. Prove Theorem 5.3.2.

33. Prove that if $\mathbf{u}$ and $\mathbf{v}$ are vectors in C^n, then

$$\mathbf{u}\cdot\mathbf{v} = \frac{1}{4}\|\mathbf{u}+\mathbf{v}\|^2 - \frac{1}{4}\|\mathbf{u}-\mathbf{v}\|^2 + \frac{i}{4}\|\mathbf{u}+i\mathbf{v}\|^2 - \frac{i}{4}\|\mathbf{u}-i\mathbf{v}\|^2$$

34. It follows from Theorem 5.3.7 that the eigenvalues of the rotation matrix

$$R_\phi = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

are $\lambda = \cos\phi \pm i\sin\phi$. Prove that if $\mathbf{x}$ is an eigenvector corresponding to either eigenvalue, then $\text{Re}(\mathbf{x})$ and $\text{Im}(\mathbf{x})$ are orthogonal and have the same length. [*Note:* This implies that $P = [\text{Re}(\mathbf{x}) \mid \text{Im}(\mathbf{x})]$ is a real scalar multiple of an orthogonal matrix.]

35. The two parts of this exercise lead you through a proof of Theorem 5.3.8.

a. For notational simplicity, let

$$M = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

and let $\mathbf{u} = \text{Re}(\mathbf{x})$ and $\mathbf{v} = \text{Im}(\mathbf{x})$, so $P = [\mathbf{u} \mid \mathbf{v}]$. Show that the relationship $A\mathbf{x} = \lambda\mathbf{x}$ implies that

$$A\mathbf{x} = (a\mathbf{u} + b\mathbf{v}) + i(-b\mathbf{u} + a\mathbf{v})$$

and then equate real and imaginary parts in this equation to show that

$$AP = [A\mathbf{u} \mid A\mathbf{v}] = [a\mathbf{u} + b\mathbf{v} \mid -b\mathbf{u} + a\mathbf{v}] = PM$$

b. Show that P is invertible, thereby completing the proof, since the result in part (a) implies that $A = PMP^{-1}$. [*Hint:* If P is not invertible, then one of its column vectors is a real scalar multiple of the other, say $\mathbf{v} = c\mathbf{u}$. Substitute this into the equations $A\mathbf{u} = a\mathbf{u} + b\mathbf{v}$ and $A\mathbf{v} = -b\mathbf{u} + a\mathbf{v}$ obtained in part (a), and show that $(1 + c^2)b\mathbf{u} = \mathbf{0}$. Finally, show that this leads to a contradiction, thereby proving that P is invertible.]

36. In this problem you will prove the complex analog of the Cauchy–Schwarz inequality.

a. Prove: If k is a complex number, and $\mathbf{u}$ and $\mathbf{v}$ are vectors in C^n, then

$$(\mathbf{u} - k\mathbf{v})\cdot(\mathbf{u} - k\mathbf{v}) = \mathbf{u}\cdot\mathbf{u} - \bar{k}(\mathbf{u}\cdot\mathbf{v}) - k\overline{(\mathbf{u}\cdot\mathbf{v})} + k\bar{k}(\mathbf{v}\cdot\mathbf{v})$$

b. Use the result in part (a) to prove that

$$0 \le \mathbf{u}\cdot\mathbf{u} - \bar{k}(\mathbf{u}\cdot\mathbf{v}) - k\overline{(\mathbf{u}\cdot\mathbf{v})} + k\bar{k}(\mathbf{v}\cdot\mathbf{v})$$

c. Take $k = (\mathbf{u}\cdot\mathbf{v})/(\mathbf{v}\cdot\mathbf{v})$ in part (b) to prove that

$$|\mathbf{u}\cdot\mathbf{v}| \le \|\mathbf{u}\|\,\|\mathbf{v}\|$$

True-False Exercises

TF. In parts **(a)–(f)** determine whether the statement is true or false, and justify your answer.

a. There is a real 5×5 matrix with no real eigenvalues.

b. The eigenvalues of a 2×2 complex matrix are the solutions of the equation $\lambda^2 - \text{tr}(A)\lambda + \det(A) = 0$.

c. A 2×2 matrix A with real entries has two distinct eigenvalues if and only if $\text{tr}(A)^2 \ne 4\det(A)$.

d. If λ is a complex eigenvalue of a real matrix A with a corresponding complex eigenvector $\mathbf{v}$, then $\bar{\lambda}$ is a complex eigenvalue of A and $\bar{\mathbf{v}}$ is a complex eigenvector of A corresponding to $\bar{\lambda}$.

e. Every eigenvalue of a complex symmetric matrix is real.

f. If a 2×2 real matrix A has complex eigenvalues and $\mathbf{x}_0$ is a vector in R^2, then the vectors $\mathbf{x}_0, A\mathbf{x}_0, A^2\mathbf{x}_0, \ldots, A^n\mathbf{x}_0, \ldots$ lie on an ellipse.

5.4 Differential Equations

Many laws of physics, chemistry, biology, engineering, and economics are described in terms of "differential equations"—that is, equations involving functions and their derivatives. In this section we will illustrate one way in which matrix diagonalization can be used to solve systems of differential equations. Calculus is a prerequisite for this section.

Terminology

Recall from calculus that a ***differential equation*** is an equation involving unknown functions and their derivatives. The ***order*** of a differential equation is the order of the highest derivative it contains. The simplest differential equations are the first-order equations of the form

$$y' = ay \tag{1}$$

where $y = f(x)$ is an unknown differentiable function to be determined, $y' = dy/dx$ is its derivative, and a is a constant. As with most differential equations, this equation has infinitely many solutions; they are the functions of the form

$$y = ce^{ax} \tag{2}$$

where c is an arbitrary constant. That every function of this form is a solution of (1) follows from the computation

$$y' = cae^{ax} = ay$$

and that these are the only solutions is shown in the exercises. Accordingly, we call (2) the ***general solution*** of (1). As an example, the general solution of the differential equation $y' = 5y$ is

$$y = ce^{5x} \tag{3}$$

Often, a physical problem that leads to a differential equation imposes some conditions that enable us to isolate one particular solution from the general solution. For example, if we require that solution (3) of the equation $y' = 5y$ satisfy the added condition

$$y(0) = 6 \tag{4}$$

(that is, $y = 6$ when $x = 0$), then on substituting these values in (3), we obtain $6 = ce^0 = c$, from which we conclude that

$$y = 6e^{5x}$$

is the only solution $y' = 5y$ that satisfies (4).

A condition such as (4), which specifies the value of the general solution at a point, is called an ***initial condition***, and the problem of solving a differential equation subject to an initial condition is called an ***initial-value problem***.

First-Order Linear Systems

In this section we will be concerned with solving systems of differential equations of the form

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n \\ y_2' &= a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n \\ &\;\;\vdots \qquad\quad \vdots \qquad\qquad\qquad\quad \vdots \\ y_n' &= a_{n1}y_1 + a_{n2}y_2 + \cdots + a_{nn}y_n \end{aligned} \tag{5}$$

where $y_1 = f_1(x),\ y_2 = f_2(x), \ldots, y_n = f_n(x)$ are functions to be determined, and the a_{ij}'s are constants. In matrix notation, (5) can be written as

$$\begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

or more briefly as

$$\boxed{\mathbf{y}' = A\mathbf{y}} \tag{6}$$

where the notation $\mathbf{y}'$ denotes the vector obtained by differentiating each component of $\mathbf{y}$.

We call (5) or its matrix form (6) a ***constant coefficient first-order homogeneous linear system***. It is of first order because all derivatives are of that order, it is linear because differentiation and matrix multiplication are linear transformations, and it is homogeneous because

$$y_1 = y_2 = \cdots = y_n = 0$$

is a solution regardless of the values of the coefficients. As expected, this is called the ***trivial solution***. In this section we will work primarily with the matrix form. Here is an example.

EXAMPLE 1 | Solution of a Linear System with Initial Conditions

(*a*) Write the following system in matrix form:

$$\begin{aligned} y_1' &= 3y_1 \\ y_2' &= -2y_2 \\ y_3' &= 5y_3 \end{aligned} \tag{7}$$

(*b*) Solve the system.

(*c*) Find a solution of the system that satisfies the initial conditions $y_1(0) = 1$, $y_2(0) = 4$, and $y_3(0) = -2$.

Solution (*a*)

$$\begin{bmatrix} y_1' \\ y_2' \\ y_3' \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \tag{8}$$

or

$$\mathbf{y}' = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{bmatrix} \mathbf{y} \tag{9}$$

Solution (*b*) Because each equation in (7) involves only one unknown function, we can solve the equations individually. It follows from (2) that these solutions are

$$\begin{aligned} y_1 &= c_1 e^{3x} \\ y_2 &= c_2 e^{-2x} \\ y_3 &= c_3 e^{5x} \end{aligned}$$

or, in matrix notation,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} c_1 e^{3x} \\ c_2 e^{-2x} \\ c_3 e^{5x} \end{bmatrix} \tag{10}$$

Solution (*c*) From the given initial conditions, we obtain

$$\begin{aligned} 1 &= y_1(0) = c_1 e^0 = c_1 \\ 4 &= y_2(0) = c_2 e^0 = c_2 \\ -2 &= y_3(0) = c_3 e^0 = c_3 \end{aligned}$$

so the solution satisfying these conditions is

$$y_1 = e^{3x}, \quad y_2 = 4e^{-2x}, \quad y_3 = -2e^{5x}$$

or, in matrix notation,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} e^{3x} \\ 4e^{-2x} \\ -2e^{5x} \end{bmatrix}$$

Solution by Diagonalization

What made the system in Example 1 easy to solve was the fact that each equation involved only one of the unknown functions, so its matrix formulation, $\mathbf{y}' = A\mathbf{y}$, had a *diagonal* coefficient matrix A [Formula (9)]. A more complicated situation occurs when some or all of the equations in the system involve more than one of the unknown functions, for in this case the coefficient matrix is not diagonal. Let us now consider how we might solve such a system.

The basic idea for solving a system $\mathbf{y}' = A\mathbf{y}$ whose coefficient matrix A is not diagonal is to introduce a new unknown vector $\mathbf{u}$ that is related to the unknown vector $\mathbf{y}$ by an equation of the form $\mathbf{y} = P\mathbf{u}$ in which P is an invertible matrix that diagonalizes A. Of course, such a matrix may or may not exist, but if it does, then we can rewrite the equation $\mathbf{y}' = A\mathbf{y}$ as

$$P\mathbf{u}' = A(P\mathbf{u})$$

or alternatively as

$$\mathbf{u}' = (P^{-1}AP)\mathbf{u}$$

Since P is assumed to diagonalize A, this equation has the form

$$\mathbf{u}' = D\mathbf{u}$$

where D is diagonal. We can now solve this equation for $\mathbf{u}$ using the method of Example 1, and then obtain $\mathbf{y}$ by matrix multiplication using the relationship $\mathbf{y} = P\mathbf{u}$.

In summary, we have the following procedure for solving a system $\mathbf{y}' = A\mathbf{y}$ in the case were A is diagonalizable.

A Procedure for Solving $\mathbf{y}' = A\mathbf{y}$ If A Is Diagonalizable

Step 1. Find a matrix P that diagonalizes A.

Step 2. Make the substitutions $\mathbf{y} = P\mathbf{u}$ and $\mathbf{y}' = P\mathbf{u}'$ to obtain a new "diagonal system" $\mathbf{u}' = D\mathbf{u}$, where $D = P^{-1}AP$.

Step 3. Solve $\mathbf{u}' = D\mathbf{u}$.

Step 4. Determine $\mathbf{y}$ from the equation $\mathbf{y} = P\mathbf{u}$.

EXAMPLE 2 | Solution Using Diagonalization

(*a*) Solve the system

$$\begin{aligned} y_1' &= y_1 + y_2 \\ y_2' &= 4y_1 - 2y_2 \end{aligned}$$

(*b*) Find the solution that satisfies the initial conditions $y_1(0) = 1$, $y_2(0) = 6$.

Solution (*a*) The coefficient matrix for the system is

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

As discussed in Section 5.2, A will be diagonalized by any matrix P whose columns are linearly independent eigenvectors of A. Since

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - 1 & -1 \\ -4 & \lambda + 2 \end{vmatrix} = \lambda^2 + \lambda - 6 = (\lambda + 3)(\lambda - 2)$$

the eigenvalues of A are $\lambda = 2$ and $\lambda = -3$. By definition,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

is an eigenvector of A corresponding to λ if and only if $\mathbf{x}$ is a nontrivial solution of

$$\begin{bmatrix} \lambda - 1 & -1 \\ -4 & \lambda + 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

If $\lambda = 2$, this system becomes

$$\begin{bmatrix} 1 & -1 \\ -4 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Solving this system yields $x_1 = t$, $x_2 = t$, so

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} t \\ t \end{bmatrix} = t\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Thus,

$$\mathbf{p}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

is a basis for the eigenspace corresponding to $\lambda = 2$. Similarly, you can show that

$$\mathbf{p}_2 = \begin{bmatrix} -\frac{1}{4} \\ 1 \end{bmatrix}$$

is a basis for the eigenspace corresponding to $\lambda = -3$. Thus,

$$P = \begin{bmatrix} 1 & -\frac{1}{4} \\ 1 & 1 \end{bmatrix}$$

diagonalizes A, and

$$D = P^{-1}AP = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}$$

Thus, as noted in Step 2 of the procedure stated above, the substitution

$$\mathbf{y} = P\mathbf{u} \quad \text{and} \quad \mathbf{y}' = P\mathbf{u}'$$

yields the "diagonal system"

$$\mathbf{u}' = D\mathbf{u} = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}\mathbf{u} \quad \text{or} \quad \begin{aligned} u_1' &= 2u_1 \\ u_2' &= -3u_2 \end{aligned}$$

From (2) the solution of this system is

$$\begin{aligned} u_1 &= c_1 e^{2x} \\ u_2 &= c_2 e^{-3x} \end{aligned} \quad \text{or} \quad \mathbf{u} = \begin{bmatrix} c_1 e^{2x} \\ c_2 e^{-3x} \end{bmatrix}$$

so the equation $\mathbf{y} = P\mathbf{u}$ yields, as the solution for $\mathbf{y}$,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{4} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} c_1 e^{2x} \\ c_2 e^{-3x} \end{bmatrix} = \begin{bmatrix} c_1 e^{2x} - \frac{1}{4} c_2 e^{-3x} \\ c_1 e^{2x} + c_2 e^{-3x} \end{bmatrix}$$

or

$$\begin{aligned} y_1 &= c_1 e^{2x} - \tfrac{1}{4} c_2 e^{-3x} \\ y_2 &= c_1 e^{2x} + c_2 e^{-3x} \end{aligned} \tag{11}$$

Solution (*b*) If we substitute the given initial conditions in (11), we obtain

$$\begin{aligned} c_1 - \tfrac{1}{4} c_2 &= 1 \\ c_1 + c_2 &= 6 \end{aligned}$$

Solving this system, we obtain $c_1 = 2$, $c_2 = 4$, so it follows from (11) that the solution satisfying the initial conditions is

$$\begin{aligned} y_1 &= 2e^{2x} - e^{-3x} \\ y_2 &= 2e^{2x} + 4e^{-3x} \end{aligned}$$

Remark Keep in mind that the method of Example 2 works because the coefficient matrix of the system is diagonalizable. In cases where this is not so, other methods are required. These are typically discussed in books devoted to differential equations.

Exercise Set 5.4

1. **a.** Solve the system

$$\begin{aligned} y_1' &= y_1 + 4y_2 \\ y_2' &= 2y_1 + 3y_2 \end{aligned}$$

b. Find the solution that satisfies the initial conditions $y_1(0) = 0$, $y_2(0) = 0$.

2. **a.** Solve the system

$$\begin{aligned} y_1' &= y_1 + 3y_2 \\ y_2' &= 4y_1 + 5y_2 \end{aligned}$$

b. Find the solution that satisfies the conditions $y_1(0) = 2$, $y_2(0) = 1$.

3. **a.** Solve the system

$$\begin{aligned} y_1' &= 4y_1 + y_3 \\ y_2' &= -2y_1 + y_2 \\ y_3' &= -2y_1 + y_3 \end{aligned}$$

b. Find the solution that satisfies the initial conditions $y_1(0) = -1$, $y_2(0) = 1$, $y_3(0) = 0$.

4. Solve the system

$$\begin{aligned} y_1' &= 4y_1 + 2y_2 + 2y_3 \\ y_2' &= 2y_1 + 4y_2 + 2y_3 \\ y_3' &= 2y_1 + 2y_2 + 4y_3 \end{aligned}$$

5. Show that every solution of $y' = ay$ has the form $y = ce^{ax}$. [*Hint:* Let $y = f(x)$ be a solution of the equation, and show that $f(x)e^{-ax}$ is constant.]

6. Show that if A is diagonalizable and

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

is a solution of the system $\mathbf{y}' = A\mathbf{y}$, then each y_i is a linear combination of $e^{\lambda_1 x}, e^{\lambda_2 x}, \ldots, e^{\lambda_n x}$, where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A.

7. Sometimes it is possible to solve a single higher-order linear differential equation with constant coefficients by expressing it as a system and applying the methods of this section. For the differential equation $y'' - y' - 6y = 0$, show that the substitutions $y_1 = y$ and $y_2 = y'$ lead to the system

$$\begin{aligned} y_1' &= y_2 \\ y_2' &= 6y_1 + y_2 \end{aligned}$$

Solve this system, and use the result to solve the original differential equation.

8. Use the procedure in Exercise 7 to solve $y'' + y' - 12y = 0$.

9. Explain how you might use the procedure in Exercise 7 to solve $y''' - 6y'' + 11y' - 6y = 0$. Use that procedure to solve the equation.

10. Solve the nondiagonalizable system

$$\begin{aligned} y_1' &= y_1 + y_2 \\ y_2' &= y_2 \end{aligned}$$

[*Hint:* Solve the second equation for y_2, substitute in the first equation, and then multiply both sides of the resulting equation by e^{-x}.]

11. Consider a system of differential equations $\mathbf{y}' = A\mathbf{y}$, where A is a 2×2 matrix. For what values of $a_{11}, a_{12}, a_{21}, a_{22}$ do the component solutions $y_1(t)$, $y_2(t)$ tend to zero as $t \to \infty$? In particular, what must be true about the determinant and the trace of A for this to happen?

12. **a.** By rewriting (11) in matrix form, show that the solution of the system in Example 2 can be expressed as

$$\mathbf{y} = c_1 e^{2x} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 e^{-3x} \begin{bmatrix} -\frac{1}{4} \\ 1 \end{bmatrix}$$

This is called the ***general solution*** of the system.

b. Note that in part (a), the vector in the first term is an eigenvector corresponding to the eigenvalue $\lambda_1 = 2$, and the vector in the second term is an eigenvector corresponding to the eigenvalue $\lambda_2 = -3$. This is a special case of the following general result:

Theorem

If the coefficient matrix A of the system $\mathbf{y}' = A\mathbf{y}$ is diagonalizable, then the general solution of the system can be expressed as

$$\mathbf{y} = c_1 e^{\lambda_1 x}\mathbf{x}_1 + c_2 e^{\lambda_2 x}\mathbf{x}_2 + \cdots + c_n e^{\lambda_n x}\mathbf{x}_n$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A, and $\mathbf{x}_i$ is an eigenvector of A corresponding to λ_i.

13. The electrical circuit in the accompanying figure is called a ***parallel LRC circuit***; it contains a resistor with resistance R ohms (Ω), an inductor with inductance L henries (H), and a capacitor with capacitance C farads (F). It is shown in electrical circuit analysis that at time t the current i_L through the inductor and the voltage v_C across the capacitor are solutions of the system

$$\begin{bmatrix} i_L'(t) \\ v_C'(t) \end{bmatrix} = \begin{bmatrix} 0 & 1/L \\ -1/C & -1/(RC) \end{bmatrix} \begin{bmatrix} i_L(t) \\ v_C(t) \end{bmatrix}$$

a. Find the general solution of this system in the case where $R = 1$ ohm, $L = 1$ henry, and $C = 0.5$ farad.

b. Find $i_L(t)$ and $v_C(t)$ subject to the initial conditions $i_L(0) = 2$ amperes and $v_C(0) = 1$ volt.

c. What can you say about the current and voltage in part (b) over the "long term" (that is, as $t \to \infty$)?

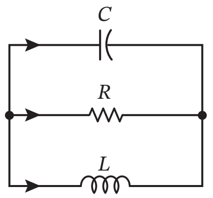


FIGURE Ex-13

In Exercises 14–15, a mapping

$$L: C^\infty(-\infty, \infty) \to C^\infty(-\infty, \infty)$$

is given.

a. Show that L is a linear operator.

b. Use the ideas in Exercises 7 and 9 to solve the differential equation $L(y) = 0$.

14. $L(y) = y'' + 2y' - 3y$

15. $L(y) = y''' - 2y'' - y' + 2y$

Working with Proofs

16. Prove the theorem in Exercise 12 by tracing through the four-step procedure preceding Example 2 with

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \quad \text{and} \quad P = [\mathbf{x}_1 \mid \mathbf{x}_2 \mid \cdots \mid \mathbf{x}_n]$$

True-False Exercises

TF. In parts (a)–(e) determine whether the statement is true or false, and justify your answer.

a. Every system of differential equations $\mathbf{y}' = A\mathbf{y}$ has a solution.

b. If $\mathbf{x}' = A\mathbf{x}$ and $\mathbf{y}' = A\mathbf{y}$, then $\mathbf{x} = \mathbf{y}$.

c. If $\mathbf{x}' = A\mathbf{x}$ and $\mathbf{y}' = A\mathbf{y}$, then $(c\mathbf{x} + d\mathbf{y})' = A(c\mathbf{x} + d\mathbf{y})$ for all scalars c and d.

d. If A is a square matrix with distinct real eigenvalues, then it is possible to solve $\mathbf{x}' = A\mathbf{x}$ by diagonalization.

e. If A and P are similar matrices, then $\mathbf{y}' = A\mathbf{y}$ and $\mathbf{u}' = P\mathbf{u}$ have the same solutions.

Working with Technology

T1. a. Find the general solution of the following system by computing appropriate eigenvalues and eigenvectors.

$$\begin{aligned} y_1' &= 3y_1 + 2y_2 + 2y_3 \\ y_2' &= y_1 + 4y_2 + y_3 \\ y_3' &= -2y_1 - 4y_2 - y_3 \end{aligned}$$

b. Find the solution that satisfies the initial conditions $y_1(0) = 0$, $y_2(0) = 1$, $y_3(0) = -3$. [*Technology not required.*]

T2. It is shown in electrical circuit theory that for the *LRC* circuit in Figure Ex-13 the current I in amperes (A) through the inductor and the voltage drop V in volts (V) across the capacitor satisfy the system of differential equations

$$\frac{dI}{dt} = \frac{V}{L}$$

$$\frac{dV}{dt} = -\frac{I}{C} - \frac{V}{RC}$$

where the derivatives are with respect to the time t. Find I and V as functions of t if $L = 0.5$ H, $C = 0.2$ F, $R = 2\ \Omega$, and the initial values of V and I are $V(0) = 1$ V and $I(0) = 2$ A.

5.5 Dynamical Systems and Markov Chains

In this optional section we will show how matrix methods can be used to analyze the behavior of physical systems that evolve over time. The methods that we will study here have been applied to problems in business, ecology, demographics, sociology, and most of the physical sciences.

Dynamical Systems

A ***dynamical system*** is a finite set of variables whose values change with time. The value of a variable at a point in time is called the ***state of the variable*** at that time, and the vector formed from these states is called the ***state vector*** (or ***state***) of the dynamical system at that time. Our primary objective in this section is to analyze how the state vector of a dynamical system changes with time. Let us begin with an example.

EXAMPLE 1 | Market Share as a Dynamical System

Suppose that two competing television channels, channel 1 and channel 2, each have 50% of the viewer market at some initial point in time. Assume that over each one-year period channel 1 captures 10% of channel 2's share, and channel 2 captures 20% of channel 1's share (see **Figure 5.5.1**). What is each channel's market share after one year?

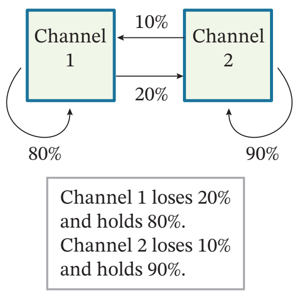


FIGURE 5.5.1

Solution Let us begin by introducing the time-dependent variables

$$\begin{aligned} x_1(t) &= \text{fraction of the market held by channel 1 at time } t \\ x_2(t) &= \text{fraction of the market held by channel 2 at time } t \end{aligned}$$

and the column vector

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \begin{array}{l} \leftarrow \textbf{Channel 1's fraction of the market at time } t \textbf{ in years} \\ \leftarrow \textbf{Channel 2's fraction of the market at time } t \textbf{ in years} \end{array}$$

The variables $x_1(t)$ and $x_2(t)$ form a dynamical system whose state at time t is the vector $\mathbf{x}(t)$. If we take $t = 0$ to be the starting point at which the two channels had 50% of the market, then the state of the system at that time is

$$\mathbf{x}(0) = \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \begin{array}{l} \leftarrow \textbf{Channel 1's fraction of the market at time } t = 0 \\ \leftarrow \textbf{Channel 2's fraction of the market at time } t = 0 \end{array} \qquad (1)$$

Now let us try to find the state of the system at time $t = 1$ (one year later). Over the one-year period, channel 1 retains 80% of its initial 50%, and it gains 10% of channel 2's initial 50%. Thus,

$$\mathbf{x}_1(1) = 0.8(0.5) + 0.1(0.5) = 0.45 \tag{2}$$

Similarly, channel 2 gains 20% of channel 1's initial 50%, and retains 90% of its initial 50%. Thus,

$$x_2(1) = 0.2(0.5) + 0.9(0.5) = 0.55 \tag{3}$$

Therefore, the state of the system at time $t = 1$ is

$$\mathbf{x}(1) = \begin{bmatrix} x_1(1) \\ x_2(1) \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix} \begin{matrix} \leftarrow \textbf{Channel 1's fraction of the market at time } t = 1 \\ \leftarrow \textbf{Channel 2's fraction of the market at time } t = 1 \end{matrix} \tag{4}$$

EXAMPLE 2 | Evolution of Market Share over Five Years

Track the market shares of channels 1 and 2 in Example 1 over a five-year period.

Solution To solve this problem suppose that we have already computed the market share of each channel at time $t = k$ and we are interested in using the known values of $x_1(k)$ and $x_2(k)$ to compute the market shares $x_1(k+1)$ and $x_2(k+1)$ one year later. The analysis is exactly the same as that used to obtain Equations (2) and (3). Over the one-year period, channel 1 retains 80% of its starting fraction $x_1(k)$ and gains 10% of channel 2's starting fraction $x_2(k)$. Thus,

$$x_1(k+1) = (0.8)x_1(k) + (0.1)x_2(k) \tag{5}$$

Similarly, channel 2 gains 20% of channel 1's starting fraction $x_1(k)$ and retains 90% of its own starting fraction $x_2(k)$. Thus,

$$x_2(k+1) = (0.2)x_1(k) + (0.9)x_2(k) \tag{6}$$

Equations (5) and (6) can be expressed in matrix form as

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \tag{7}$$

which provides a way of using matrix multiplication to compute the state of the system at time $t = k + 1$ from the state at time $t = k$. For example, using (1) and (7) we obtain

$$\mathbf{x}(1) = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \mathbf{x}(0) = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix}$$

which agrees with (4). Similarly,

$$\mathbf{x}(2) = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \mathbf{x}(1) = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix} = \begin{bmatrix} 0.415 \\ 0.585 \end{bmatrix}$$

We can now continue this process, using Formula (7) to compute $\mathbf{x}(3)$ from $\mathbf{x}(2)$, then $\mathbf{x}(4)$ from $\mathbf{x}(3)$, and so on. This yields (verify)

$$\mathbf{x}(3) = \begin{bmatrix} 0.3905 \\ 0.6095 \end{bmatrix}, \quad \mathbf{x}(4) = \begin{bmatrix} 0.37335 \\ 0.62665 \end{bmatrix}, \quad \mathbf{x}(5) = \begin{bmatrix} 0.361345 \\ 0.638655 \end{bmatrix} \tag{8}$$

Thus, after five years, channel 1 will hold about 36% of the market and channel 2 will hold about 64% of the market.

If desired, we can continue the market analysis in the last example beyond the five-year period and explore what happens to the market share over the long term. We did so, using a computer, and obtained the following state vectors (rounded to six decimal places):

$$\mathbf{x}(10) \approx \begin{bmatrix} 0.338041 \\ 0.661959 \end{bmatrix}, \quad \mathbf{x}(20) \approx \begin{bmatrix} 0.333466 \\ 0.666534 \end{bmatrix}, \quad \mathbf{x}(40) \approx \begin{bmatrix} 0.333333 \\ 0.666667 \end{bmatrix} \tag{9}$$

All subsequent state vectors, when rounded to six decimal places, are the same as $\mathbf{x}(40)$, so we see that the market shares eventually stabilize with channel 1 holding about one-third of the market and channel 2 holding about two-thirds. Later in this section, we will explain why this stabilization occurs.

Markov Chains

In many dynamical systems the states of the variables are not known with certainty but can be expressed as probabilities; such dynamical systems are called ***stochastic processes*** (from the Greek word *stochastikos*, meaning "proceeding by guesswork"). A detailed study of stochastic processes requires a precise definition of the term *probability*, which is outside the scope of this course. However, the following interpretation will suffice for our present purposes:

> *Stated informally, the **probability** that an experiment or observation will have a certain outcome is the fraction of the time that the outcome would occur if the experiment could be repeated indefinitely under constant conditions—the greater the number of actual repetitions, the more accurately the probability describes the fraction of time that the outcome occurs.*

For example, when we say that the probability of tossing heads with a fair coin is $\frac{1}{2}$, we mean that if the coin were tossed many times under constant conditions, then we would expect about half of the outcomes to be heads. Probabilities are often expressed as decimals or percentages. Thus, the probability of tossing heads with a fair coin can also be expressed as 0.5 or 50%.

If an experiment or observation has n possible outcomes, then the probabilities of those outcomes must be nonnegative fractions whose sum is 1. The probabilities are nonnegative because each describes the fraction of occurrences of an outcome over the long term, and the sum is 1 because they account for all possible outcomes. For example, if a box containing 10 balls has one red ball, three green balls, and six yellow balls, and if a ball is drawn at random from the box, then the probabilities of the various outcomes are

$$\begin{aligned} p_1 &= \text{prob(red)} = 1/10 = 0.1 \\ p_2 &= \text{prob(green)} = 3/10 = 0.3 \\ p_3 &= \text{prob(yellow)} = 6/10 = 0.6 \end{aligned}$$

Each probability is a nonnegative fraction and

$$p_1 + p_2 + p_3 = 0.1 + 0.3 + 0.6 = 1$$

In a stochastic process with n possible states, the state vector at each time t has the form

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} \begin{array}{l} \textbf{Probability that the system is in state 1} \\ \textbf{Probability that the system is in state 2} \\ \vdots \\ \textbf{Probability that the system is in state } \boldsymbol{n} \end{array}$$

The entries in this vector must add up to 1 since they account for all n possibilities. In general, a vector with nonnegative entries that add up to 1 is called a ***probability vector***.

EXAMPLE 3 | Example 1 Revisited from the Probability Viewpoint

Observe that the state vectors in Examples 1 and 2 are all probability vectors. This is to be expected since the entries in each state vector are the fractional market shares of the channels, and together they account for the entire market. In practice, it is preferable to interpret the entries in the state vectors as probabilities rather than exact market fractions, since market information is usually obtained by statistical sampling procedures with intrinsic uncertainties. Thus, for example, the state vector

$$\mathbf{x}(1) = \begin{bmatrix} x_1(1) \\ x_2(1) \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix}$$

which we interpreted in Example 1 to mean that channel 1 has 45% of the market and channel 2 has 55%, can also be interpreted to mean that an individual picked at random from the market will be a channel 1 viewer with probability 0.45 and a channel 2 viewer with probability 0.55.

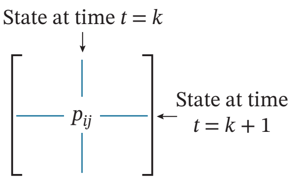
A square matrix whose columns are probability vectors is called a ***stochastic matrix***. Such matrices commonly occur in formulas that relate successive states of a stochastic process. For example, the state vectors $\mathbf{x}(k+1)$ and $\mathbf{x}(k)$ in (7) are related by an equation of the form $\mathbf{x}(k+1) = P\mathbf{x}(k)$ in which

$$P = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \tag{10}$$

is a stochastic matrix. It should not be surprising that the column vectors of P are probability vectors, since the entries in each column provide a breakdown of what happens to each channel's market share over the year—the entries in column 1 convey that each year channel 1 retains 80% of its market share and loses 20%; and the entries in column 2 convey that each year channel 2 retains 90% of its market share and loses 10%. The entries in (10) can also be viewed as probabilities:

$$\begin{aligned} p_{11} &= 0.8 = \text{probability that a channel 1 viewer remains a channel 1 viewer} \\ p_{21} &= 0.2 = \text{probability that a channel 1 viewer becomes a channel 2 viewer} \\ p_{12} &= 0.1 = \text{probability that a channel 2 viewer becomes a channel 1 viewer} \\ p_{22} &= 0.9 = \text{probability that a channel 2 viewer remains a channel 2 viewer} \end{aligned}$$

Example 1 is a special case of a large class of stochastic processes called *Markov chains*.



The entry p_{ij} is the probability that the system is in state i at time $t = k + 1$ if it is in state j at time $t = k$.

FIGURE 5.5.2

Definition 1

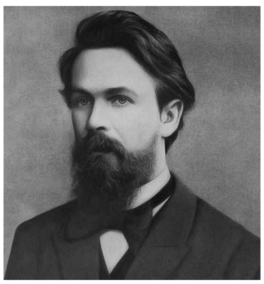
A ***Markov chain*** is a dynamical system whose state vectors at a succession of equally spaced times are probability vectors and for which the state vectors at successive times are related by an equation of the form

$$\mathbf{x}(k+1) = P\mathbf{x}(k)$$

in which $P = [p_{ij}]$ is a stochastic matrix and p_{ij} is the probability that the system will be in state i at time $t = k + 1$ if it is in state j at time $t = k$. The matrix P is called the ***transition matrix*** for the system.

Warning **Note that in this definition the row index *i* corresponds to the later state and the column index *j* to the earlier state (Figure 5.5.2).**

Historical Note



Andrei Andreyevich Markov (1856–1922)

Markov chains are named in honor of the Russian mathematician A. A. Markov, a lover of poetry, who used them to analyze the alternation of vowels and consonants in the poem *Eugene Onegin* by Pushkin. Markov believed that the only applications of his chains were to the analysis of literary works, so he would be astonished to learn that his discovery is used today in the social sciences, quantum theory, and genetics!

[Image: https://en.wikipedia.org/wiki/Andrey_Markov#/media/File:Andrei_Markov.jpg. Public domain.]

EXAMPLE 4 | Wildlife Migration as a Markov Chain

Suppose that a tagged lion can migrate over three adjacent game reserves in search of food: Reserve 1, Reserve 2, and Reserve 3. Based on data about the food resources, researchers conclude that the monthly migration pattern of the lion can be modeled by a Markov chain with transition matrix

$$\begin{array}{c} \text{Reserve at time } t = k \\ \begin{array}{ccc} 1 & 2 & 3 \end{array} \end{array}$$

$$P = \begin{bmatrix} 0.5 & 0.4 & 0.6 \\ 0.2 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.1 \end{bmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \quad \text{Reserve at time } t = k+1$$

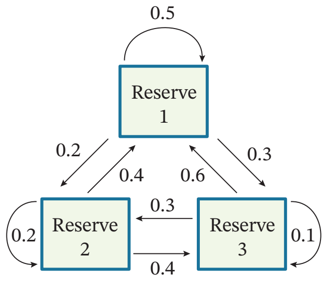


FIGURE 5.5.3

(see **Figure 5.5.3**). That is,

$p_{11} = 0.5 =$ probability that the lion will stay in Reserve 1 when it is in Reserve 1
$p_{12} = 0.4 =$ probability that the lion will move from Reserve 2 to Reserve 1
$p_{13} = 0.6 =$ probability that the lion will move from Reserve 3 to Reserve 1
$p_{21} = 0.2 =$ probability that the lion will move from Reserve 1 to Reserve 2
$p_{22} = 0.2 =$ probability that the lion will stay in Reserve 2 when it is in Reserve 2
$p_{23} = 0.3 =$ probability that the lion will move from Reserve 3 to Reserve 2
$p_{31} = 0.3 =$ probability that the lion will move from Reserve 1 to Reserve 3
$p_{32} = 0.4 =$ probability that the lion will move from Reserve 2 to Reserve 3
$p_{33} = 0.1 =$ probability that the lion will stay in Reserve 3 when it is in Reserve 3

Assuming that t is in months and the lion is released in Reserve 2 at time $t = 0$, track its probable locations over a six-month period, and find the reserve in which it is most likely to be at the end of that period.

Solution Let $x_1(k)$, $x_2(k)$, and $x_3(k)$ be the probabilities that the lion is in Reserve 1, 2, or 3, respectively, at time $t = k$, and let

$$\mathbf{x}(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix}$$

be the state vector at that time. Since we know with certainty that the lion is in Reserve 2 at time $t = 0$, the initial state vector is

$$\mathbf{x}(0) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

We leave it for you to use a calculator or computer to show that the state vectors over a six-month period are

$$\mathbf{x}(1) = P\mathbf{x}(0) = \begin{bmatrix} 0.400 \\ 0.200 \\ 0.400 \end{bmatrix}, \quad \mathbf{x}(2) = P\mathbf{x}(1) = \begin{bmatrix} 0.520 \\ 0.240 \\ 0.240 \end{bmatrix}, \quad \mathbf{x}(3) = P\mathbf{x}(2) = \begin{bmatrix} 0.500 \\ 0.224 \\ 0.276 \end{bmatrix}$$

$$\mathbf{x}(4) = P\mathbf{x}(3) \approx \begin{bmatrix} 0.505 \\ 0.228 \\ 0.267 \end{bmatrix}, \quad \mathbf{x}(5) = P\mathbf{x}(4) \approx \begin{bmatrix} 0.504 \\ 0.227 \\ 0.269 \end{bmatrix}, \quad \mathbf{x}(6) = P\mathbf{x}(5) \approx \begin{bmatrix} 0.504 \\ 0.227 \\ 0.269 \end{bmatrix}$$

As in Example 2, the state vectors here seem to stabilize over time with a probability of approximately 0.504 that the lion is in Reserve 1, a probability of approximately 0.227 that it is in Reserve 2, and a probability of approximately 0.269 that it is in Reserve 3.

From $\mathbf{x}(6)$ we see that the lion is most likely to be in Reserve 1 at the end of six months.

Markov Chains in Terms of Powers of the Transition Matrix

In a Markov chain with an initial state of $\mathbf{x}(0)$, the successive state vectors are

$$\boxed{\mathbf{x}(1) = P\mathbf{x}(0), \quad \mathbf{x}(2) = P\mathbf{x}(1), \quad \mathbf{x}(3) = P\mathbf{x}(2), \quad \mathbf{x}(4) = P\mathbf{x}(3), \ldots}$$

For brevity, it is common to denote $\mathbf{x}(k)$ by $\mathbf{x}_k$, which allows us to write the successive state vectors more briefly as

$$\boxed{\mathbf{x}_1 = P\mathbf{x}_0, \quad \mathbf{x}_2 = P\mathbf{x}_1, \quad \mathbf{x}_3 = P\mathbf{x}_2, \quad \mathbf{x}_4 = P\mathbf{x}_3, \ldots} \tag{11}$$

Note that Formula (12) makes it possible to compute any state vector without first computing the earlier state vectors as required in Formula (11).

Alternatively, these state vectors can be expressed in terms of the initial state vector $\mathbf{x}_0$ as

$$\mathbf{x}_1 = P\mathbf{x}_0, \quad \mathbf{x}_2 = P(P\mathbf{x}_0) = P^2\mathbf{x}_0, \quad \mathbf{x}_3 = P(P^2\mathbf{x}_0) = P^3\mathbf{x}_0, \quad \mathbf{x}_4 = P(P^3\mathbf{x}_0) = P^4\mathbf{x}_0, \ldots$$

from which it follows that

$$\boxed{\mathbf{x}_k = P^k\mathbf{x}_0} \tag{12}$$

EXAMPLE 5 | Finding a State Vector Directly

Use Formula (12) to find the state vector $\mathbf{x}(3)$ in Example 2.

Solution From (1) and (7), the initial state vector and transition matrix are

$$\mathbf{x}_0 = \mathbf{x}(0) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix}$$

We leave it for you to calculate P^3 and show that

$$\mathbf{x}(3) = \mathbf{x}_3 = P^3\mathbf{x}_0 = \begin{bmatrix} 0.562 & 0.219 \\ 0.438 & 0.781 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.3905 \\ 0.6095 \end{bmatrix}$$

which agrees with the result in (8).

Long-Term Behavior of a Markov Chain

We have seen two examples of Markov chains in which the state vectors seem to stabilize after a period of time. Thus, it is reasonable to ask whether all Markov chains have this property. The following example shows that this is not the case.

EXAMPLE 6 | A Markov Chain That Does Not Stabilize

The matrix

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

is stochastic and hence can be regarded as the transition matrix for a Markov chain. A simple calculation shows that $P^2 = I$, from which it follows that

$$I = P^2 = P^4 = P^6 = \cdots \quad \text{and} \quad P = P^3 = P^5 = P^7 = \cdots$$

Thus, the successive states in the Markov chain with initial vector $\mathbf{x}_0$ are

$$\mathbf{x}_0, \quad P\mathbf{x}_0, \quad \mathbf{x}_0, \quad P\mathbf{x}_0, \quad \mathbf{x}_0, \ldots$$

which oscillate between $\mathbf{x}_0$ and $P\mathbf{x}_0$. Thus, the Markov chain does not stabilize unless both components of $\mathbf{x}_0$ are $\frac{1}{2}$ (verify).

A precise definition of what it means for a sequence of numbers or vectors to stabilize is given in calculus; however, that level of precision will not be needed here. Stated informally, we will say that a sequence of vectors

$$\mathbf{x}_1, \quad \mathbf{x}_2, \ldots, \quad \mathbf{x}_k, \ldots$$

approaches a ***limit*** $\mathbf{q}$ or that it ***converges*** to $\mathbf{q}$ if all entries in $\mathbf{x}_k$ can be made as close as we like to the corresponding entries in the vector $\mathbf{q}$ by taking k sufficiently large. We denote this by writing $\mathbf{x}_k \to \mathbf{q}$ as $k \to \infty$. Similarly, we say that a sequence of matrices

$$P_1, P_2, P_3, \ldots, P_k, \ldots$$

converges to a matrix Q, written $P_k \to Q$ as $k \to \infty$, if each entry of P_k can be made as close as we like to the corresponding entry of Q by taking k sufficiently large.

We saw in Example 6 that the state vectors of a Markov chain need not approach a limit in all cases. However, by imposing a mild condition on the transition matrix of a Markov chain, we can guarantee that the state vectors will approach a limit.

Definition 2

A stochastic matrix P is said to be ***regular*** if P or some positive power of P has all positive entries, and a Markov chain whose transition matrix is regular is said to be a ***regular Markov chain***.

EXAMPLE 7 | Regular Stochastic Matrices

The transition matrices in Examples 2 and 4 are regular because their entries are positive. The matrix

$$P = \begin{bmatrix} 0.5 & 1 \\ 0.5 & 0 \end{bmatrix}$$

is regular because

$$P^2 = \begin{bmatrix} 0.75 & 0.5 \\ 0.25 & 0.5 \end{bmatrix}$$

has positive entries. The matrix P in Example 6 is not regular because P and every positive power of P have some zero entries (verify).

The following theorem, which we state without proof, is the fundamental result about the long-term behavior of Markov chains.

Theorem 5.5.1

If P is the transition matrix for a regular Markov chain, then:

(a) There is a unique probability vector $\mathbf{q}$ with positive entries such that $P\mathbf{q} = \mathbf{q}$.

(b) For any initial probability vector $\mathbf{x}_0$, the sequence of state vectors

$$\mathbf{x}_0, \quad P\mathbf{x}_0, \ldots, \quad P^k\mathbf{x}_0, \ldots$$

converges to $\mathbf{q}$.

(c) The sequence $P, P^2, \ldots, P^k, \ldots$ converges to the matrix Q each of whose column vectors is $\mathbf{q}$.

The vector $\mathbf{q}$ in Theorem 5.5.1 is called the ***steady-state*** vector of the Markov chain. Because it is a nonzero vector that satisfies the equation $P\mathbf{q} = \mathbf{q}$, it is an eigenvector corresponding to the eigenvalue $\lambda = 1$ of P. Thus, $\mathbf{q}$ can be found by solving the linear system

$$(I - P)\mathbf{q} = \mathbf{0} \tag{13}$$

subject to the requirement that $\mathbf{q}$ be a probability vector. Here are some examples.

EXAMPLE 8 | Examples 1 and 2 Revisited

The transition matrix for the Markov chain in Example 2 is

$$P = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix}$$

Since the entries of P are positive, the Markov chain is regular and hence has a unique steady-state vector $\mathbf{q}$. To find $\mathbf{q}$ we will solve the system $(I - P)\mathbf{q} = \mathbf{0}$, which we can write as

$$\begin{bmatrix} 0.2 & -0.1 \\ -0.2 & 0.1 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The general solution of this system is

$$q_1 = 0.5s, \quad q_2 = s$$

(verify), which we can write in vector form as

$$\mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} 0.5s \\ s \end{bmatrix} = \begin{bmatrix} \frac{1}{2}s \\ s \end{bmatrix} \tag{14}$$

For $\mathbf{q}$ to be a probability vector, we must have

$$1 = q_1 + q_2 = \tfrac{3}{2}s$$

which implies that $s = \frac{2}{3}$. Substituting this value in (14) yields the steady-state vector

$$\mathbf{q} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \end{bmatrix}$$

which is consistent with the numerical results obtained in (9).

EXAMPLE 9 | Example 4 Revisited

The transition matrix for the Markov chain in Example 4 is

$$P = \begin{bmatrix} 0.5 & 0.4 & 0.6 \\ 0.2 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.1 \end{bmatrix}$$

Since the entries of P are positive, the Markov chain is regular and hence has a unique steady-state vector $\mathbf{q}$. To find $\mathbf{q}$ we will solve the system $(I - P)\mathbf{q} = \mathbf{0}$, which we can write (using fractions) as

$$\begin{bmatrix} \frac{1}{2} & -\frac{2}{5} & -\frac{3}{5} \\ -\frac{1}{5} & \frac{4}{5} & -\frac{3}{10} \\ -\frac{3}{10} & -\frac{2}{5} & \frac{9}{10} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{15}$$

(We have converted to fractions to avoid roundoff error in this illustrative example.) We leave it for you to confirm that the reduced row echelon form of the coefficient matrix is

$$\begin{bmatrix} 1 & 0 & -\frac{15}{8} \\ 0 & 1 & -\frac{27}{32} \\ 0 & 0 & 0 \end{bmatrix}$$

and that the general solution of (15) is

$$q_1 = \tfrac{15}{8}s, \quad q_2 = \tfrac{27}{32}s, \quad q_3 = s \tag{16}$$

For $\mathbf{q}$ to be a probability vector we must have $q_1 + q_2 + q_3 = 1$, from which it follows that $s = \frac{32}{119}$ (verify). Substituting this value in (16) yields the steady-state vector

$$\mathbf{q} = \begin{bmatrix} \frac{60}{119} \\ \frac{27}{119} \\ \frac{32}{119} \end{bmatrix} \approx \begin{bmatrix} 0.5042 \\ 0.2269 \\ 0.2689 \end{bmatrix}$$

(verify), which is consistent with the results obtained in Example 4.

Exercise Set 5.5

In Exercises 1–2, determine whether A is a stochastic matrix. If A is not stochastic, then explain why not.

1. a. $A = \begin{bmatrix} 0.4 & 0.3 \\ 0.6 & 0.7 \end{bmatrix}$ b. $A = \begin{bmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{bmatrix}$

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

2. a. $A = \begin{bmatrix} 0.2 & 0.9 \\ 0.8 & 0.1 \end{bmatrix}$ b. $A = \begin{bmatrix} 0.2 & 0.8 \\ 0.9 & 0.1 \end{bmatrix}$

c. $A = \begin{bmatrix} \frac{1}{12} & \frac{1}{9} & \frac{1}{6} \\ \frac{1}{2} & 0 & \frac{5}{6} \\ \frac{5}{12} & \frac{8}{9} & 0 \end{bmatrix}$ d. $A = \begin{bmatrix} -1 & \frac{1}{3} & \frac{1}{2} \\ 0 & \frac{1}{3} & \frac{1}{2} \\ 2 & \frac{1}{3} & 0 \end{bmatrix}$

In Exercises 3–4, use Formulas (11) and (12) to compute the state vector $\mathbf{x}_4$ in two different ways.

3. $P = \begin{bmatrix} 0.5 & 0.6 \\ 0.5 & 0.4 \end{bmatrix}$; $\mathbf{x}_0 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$

4. $P = \begin{bmatrix} 0.8 & 0.5 \\ 0.2 & 0.5 \end{bmatrix}$; $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

In Exercises 5–6, determine whether P is a regular stochastic matrix.

5. a. $P = \begin{bmatrix} \frac{1}{5} & \frac{1}{7} \\ \frac{4}{5} & \frac{6}{7} \end{bmatrix}$ b. $P = \begin{bmatrix} \frac{1}{5} & 0 \\ \frac{4}{5} & 1 \end{bmatrix}$ c. $P = \begin{bmatrix} \frac{1}{5} & 1 \\ \frac{4}{5} & 0 \end{bmatrix}$

6. a. $P = \begin{bmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 0 \end{bmatrix}$ b. $P = \begin{bmatrix} 1 & \frac{2}{3} \\ 0 & \frac{1}{3} \end{bmatrix}$ c. $P = \begin{bmatrix} \frac{3}{4} & \frac{1}{3} \\ \frac{1}{4} & \frac{2}{3} \end{bmatrix}$

In Exercises 7–10, verify that P is a regular stochastic matrix, and find the steady-state vector for the associated Markov chain.

7. $P = \begin{bmatrix} \frac{1}{4} & \frac{2}{3} \\ \frac{3}{4} & \frac{1}{3} \end{bmatrix}$ 8. $P = \begin{bmatrix} 0.2 & 0.6 \\ 0.8 & 0.4 \end{bmatrix}$

9. $P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{4} & 0 & \frac{2}{3} \end{bmatrix}$ 10. $P = \begin{bmatrix} \frac{1}{3} & \frac{1}{4} & \frac{2}{5} \\ 0 & \frac{3}{4} & \frac{2}{5} \\ \frac{2}{3} & 0 & \frac{1}{5} \end{bmatrix}$

11. Consider a Markov process with transition matrix

$$\begin{array}{r c} & \begin{array}{cc} \textbf{State 1} & \textbf{State 2} \end{array} \\ \begin{array}{r} \textbf{State 1} \\ \textbf{State 2} \end{array} & \begin{bmatrix} 0.2 & 0.1 \\ 0.8 & 0.9 \end{bmatrix} \end{array}$$

a. What does the entry 0.2 represent?

b. What does the entry 0.1 represent?

c. If the system is in state 1 initially, what is the probability that it will be in state 2 at the next observation?

d. If the system has a 50% chance of being in state 1 initially, what is the probability that it will be in state 2 at the next observation?

12. Consider a Markov process with transition matrix

$$\begin{array}{r c} & \begin{array}{cc} \textbf{State 1} & \textbf{State 2} \end{array} \\ \begin{array}{r} \textbf{State 1} \\ \textbf{State 2} \end{array} & \begin{bmatrix} 0 & \frac{1}{7} \\ 1 & \frac{6}{7} \end{bmatrix} \end{array}$$

a. What does the entry $\frac{6}{7}$ represent?

b. What does the entry 0 represent?

c. If the system is in state 1 initially, what is the probability that it will be in state 1 at the next observation?

d. If the system has a 50% chance of being in state 1 initially, what is the probability that it will be in state 2 at the next observation?

13. On a given day the air quality in a certain city is either good or bad. Records show that when the air quality is good on one day, then there is a 95% chance that it will be good the next day, and when the air quality is bad on one day, then there is a 45% chance that it will be bad the next day.

a. Find a transition matrix for this phenomenon.

b. If the air quality is good today, what is the probability that it will be good two days from now?

c. If the air quality is bad today, what is the probability that it will be bad three days from now?

d. If there is a 20% chance that the air quality will be good today, what is the probability that it will be good tomorrow?

14. In a laboratory experiment, a mouse can choose one of two food types each day, type I or type II. Records show that if the mouse chooses type I on a given day, then there is a 75% chance that it will choose type I the next day, and if it chooses type II on one day, then there is a 50% chance that it will choose type II the next day.

a. Find a transition matrix for this phenomenon.

b. If the mouse chooses type I today, what is the probability that it will choose type I two days from now?

c. If the mouse chooses type II today, what is the probability that it will choose type II three days from now?

d. If there is a 10% chance that the mouse will choose type I today, what is the probability that it will choose type I tomorrow?

15. Suppose that at some initial point in time 100,000 people live in a certain city and 25,000 people live in its suburbs. The Regional Planning Commission determines that each year 5% of the city population moves to the suburbs and 3% of the suburban population moves to the city.

a. Assuming that the total population remains constant, make a table that shows the populations of the city and its suburbs over a five-year period (round to the nearest integer).

b. Over the long term, how will the population be distributed between the city and its suburbs?

16. Suppose that two competing television stations, station 1 and station 2, each have 50% of the viewer market at some initial point in time. Assume that over each one-year period station 1 captures 5% of station 2's market share and station 2 captures 10% of station 1's market share.

a. Make a table that shows the market share of each station over a five-year period.

b. Over the long term, how will the market share be distributed between the two stations?

17. Fill in the missing entries of the stochastic matrix

$$P = \begin{bmatrix} \frac{7}{10} & * & \frac{1}{5} \\ * & \frac{3}{10} & * \\ \frac{1}{10} & \frac{3}{5} & \frac{3}{10} \end{bmatrix}$$

and find its steady-state vector.

18. If P is an $n \times n$ stochastic matrix, and if M is a $1 \times n$ matrix whose entries are all 1's, then $MP =$ ________.

19. If P is a regular stochastic matrix with steady-state vector $\mathbf{q}$, what can you say about the sequence of products

$$P\mathbf{q}, \quad P^2\mathbf{q}, \quad P^3\mathbf{q}, \ldots, \quad P^k\mathbf{q}, \ldots$$

as $k \to \infty$?

20. a. If P is a regular $n \times n$ stochastic matrix with steady-state vector $\mathbf{q}$, and if $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ are the standard unit vectors in column form, what can you say about the behavior of the sequence

$$P\mathbf{e}_i, \quad P^2\mathbf{e}_i, \quad P^3\mathbf{e}_i, \ldots, \quad P^k\mathbf{e}_i, \ldots$$

as $k \to \infty$ for each $i = 1, 2, \ldots, n$?

b. What does this tell you about the behavior of the column vectors of P^k as $k \to \infty$?

Working with Proofs

21. Prove that the product of two stochastic matrices with the same size is a stochastic matrix. [*Hint:* Write each column of the product as a linear combination of the columns of the first factor.]

22. Prove that if P is a stochastic matrix whose entries are all greater than or equal to ρ, then the entries of P^2 are greater than or equal to ρ.

True-False Exercises

TF. In parts (a)–(g) determine whether the statement is true or false, and justify your answer.

a. The vector $\begin{bmatrix} \frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix}$ is a probability vector.

b. The matrix $\begin{bmatrix} 0.2 & 1 \\ 0.8 & 0 \end{bmatrix}$ is a regular stochastic matrix.

c. The column vectors of a transition matrix are probability vectors.

d. A steady-state vector for a Markov chain with transition matrix P is any solution of the linear system $(I - P)\mathbf{q} = \mathbf{0}$.

e. The square of every regular stochastic matrix is stochastic.

f. A vector with real entries that sum to 1 is a probability vector.

g. Every regular stochastic matrix has $\lambda = 1$ as an eigenvalue.

Working with Technology

T1. In Examples 4 and 9 we considered the Markov chain with transition matrix P and initial state vector $\mathbf{x}(0)$ where

$$P = \begin{bmatrix} 0.5 & 0.4 & 0.6 \\ 0.2 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.1 \end{bmatrix} \quad \text{and} \quad \mathbf{x}(0) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

a. Confirm the numerical values of $\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(6)$ obtained in Example 4 using the method given in that example.

b. As guaranteed by part (*c*) of Theorem 5.5.1, confirm that the sequence $P, P^2, \ldots, P^k, \ldots$ converges to the matrix Q each of whose column vectors is the steady-state vector $\mathbf{q}$ obtained in Example 9.

T2. Suppose that a car rental agency has three locations, numbered 1, 2, and 3. A customer may rent a car from any of the three locations and return it to any of the three locations. Records show that cars are rented and returned in accordance with the following probabilities:

		Rented from Location		
		1	**2**	**3**
Returned to Location	**1**	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{3}{5}$
	2	$\frac{4}{5}$	$\frac{3}{10}$	$\frac{1}{5}$
	3	$\frac{1}{10}$	$\frac{1}{2}$	$\frac{1}{5}$

a. Assuming that a car is rented from location 1, what is the probability that it will be at location 1 after two rentals?

b. Assuming that this dynamical system can be modeled as a Markov chain, find the steady-state vector.

c. If the rental agency owns 120 cars, how many parking spaces should it allocate at each location to be reasonably certain that it will have enough spaces for the cars over the long term? Explain your reasoning.

T3. Physical traits are determined by the genes that an offspring receives from its parents. In the simplest case a trait in the offspring is determined by one pair of genes, one member of the pair inherited from the male parent and the other from the female parent. Typically, each gene in a pair can assume one of two forms, called ***alleles***, denoted by A and a. This leads to three possible pairings:

$$AA, \quad Aa, \quad aa$$

called ***genotypes*** (the pairs Aa and aA determine the same trait and hence are not distinguished from one another). It is shown in the study of heredity that if a parent of known genotype is crossed with a random parent of unknown genotype, then the offspring will have the genotype probabilities given in the following table, which can be viewed as a transition matrix for a Markov process:

		Genotype of Parent		
		AA	***Aa***	***aa***
Genotype of Offspring	***AA***	$\frac{1}{2}$	$\frac{1}{4}$	0
	Aa	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
	aa	0	$\frac{1}{4}$	$\frac{1}{2}$

Thus, for example, the offspring of a parent of genotype AA that is crossed at random with a parent of unknown genotype will have a 50% chance of being AA, a 50% chance of being Aa, and no chance of being aa.

a. Show that the transition matrix is regular.

b. Find the steady-state vector and discuss its physical interpretation.

Chapter 5 Supplementary Exercises

1. **a.** Show that if $0 < \theta < \pi$, then

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

has no real eigenvalues and consequently no real eigenvectors.

b. Give a geometric explanation of the result in part (a).

2. Find the eigenvalues of

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

3. **a.** Show that if D is a diagonal matrix with nonnegative entries on the main diagonal, then there is a matrix S such that $S^2 = D$.

b. Show that if A is a diagonalizable matrix with nonnegative eigenvalues, then there is a matrix S such that $S^2 = A$.

c. Find a matrix S such that $S^2 = A$, given that

$$A = \begin{bmatrix} 1 & 3 & 1 \\ 0 & 4 & 5 \\ 0 & 0 & 9 \end{bmatrix}$$

4. Given that A and B are similar matrices, in each part determine whether the given matrices are also similar.

a. A^T and B^T

b. A^k and B^k (k is a positive integer)

c. A^{-1} and B^{-1} (if A is invertible)

5. Prove: If A is a square matrix and $p(\lambda) = \det(\lambda I - A)$ is the characteristic polynomial of A, then the coefficient of λ^{n-1} in $p(\lambda)$ is the negative of the trace of A.

6. Prove: If $b \neq 0$, then

$$A = \begin{bmatrix} a & b \\ 0 & a \end{bmatrix}$$

is not diagonalizable.

7. In advanced linear algebra, one proves the ***Cayley–Hamilton Theorem***, which states that a square matrix A satisfies its characteristic equation; that is, if

$$c_0 + c_1\lambda + c_2\lambda^2 + \cdots + c_{n-1}\lambda^{n-1} + \lambda^n = 0$$

is the characteristic equation of A, then

$$c_0 I + c_1 A + c_2 A^2 + \cdots + c_{n-1}A^{n-1} + A^n = 0$$

Verify this result for

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

In Exercises 8–10, use the Cayley–Hamilton Theorem, stated in Exercise 7.

8. **a.** Use Exercise 28 of Section 5.1 to establish the Cayley–Hamilton Theorem for 2×2 matrices.

b. Prove the Cayley–Hamilton Theorem for $n \times n$ diagonalizable matrices.

9. The Cayley–Hamilton Theorem provides a method for calculating powers of a matrix. For example, if A is a 2×2 matrix with characteristic equation

$$c_0 + c_1\lambda + \lambda^2 = 0$$

then $c_0 I + c_1 A + A^2 = 0$, so

$$A^2 = -c_1 A - c_0 I$$

Multiplying through by A yields $A^3 = -c_1 A^2 - c_0 A$, which expresses A^3 in terms of A^2 and A, and multiplying through by A^2 yields $A^4 = -c_1 A^3 - c_0 A^2$, which expresses A^4 in terms of A^3 and A^2. Continuing in this way, we can calculate successive powers of A by expressing them in terms of lower powers. Use this procedure to calculate A^2, A^3, A^4, and A^5 for

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

10. Use the method of the preceding exercise to calculate A^3 and A^4 for

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

11. Find the eigenvalues of the matrix

$$A = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ c_1 & c_2 & \cdots & c_n \\ \vdots & \vdots & & \vdots \\ c_1 & c_2 & \cdots & c_n \end{bmatrix}$$

12. **a.** It was shown in Exercise 37 of Section 5.1 that if A is an $n \times n$ matrix, then the coefficient of λ^n in the characteristic polynomial of A is 1. (A polynomial with this property is called ***monic***.) Show that the matrix

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & 0 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -c_{n-1} \end{bmatrix}$$

has characteristic polynomial

$$p(\lambda) = c_0 + c_1\lambda + \cdots + c_{n-1}\lambda^{n-1} + \lambda^n$$

This shows that every monic polynomial is the characteristic polynomial of some matrix. The matrix in this example is called the ***companion matrix*** of $p(\lambda)$. [*Hint:* Evaluate all determinants in the problem by adding a multiple of the second row to the first to introduce a zero at the top of the first column, and then expanding by cofactors along the first column.]

b. Find a matrix with characteristic polynomial

$$p(\lambda) = 1 - 2\lambda + \lambda^2 + 3\lambda^3 + \lambda^4$$

13. A square matrix A is called ***nilpotent*** if $A^n = 0$ for some positive integer n. What can you say about the eigenvalues of a nilpotent matrix?

14. Prove: If A is an $n \times n$ matrix with real entries and n is odd, then A has at least one real eigenvalue.

15. Find a 3×3 matrix A that has eigenvalues $\lambda = 0, 1$, and -1 with corresponding eigenvectors

$$\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

respectively.

16. Suppose that a 4×4 matrix A has eigenvalues $\lambda_1 = 1$, $\lambda_2 = -2$, $\lambda_3 = 3$, and $\lambda_4 = -3$.

a. Use the method of Exercise 24 of Section 5.1 to find $\det(A)$.

b. Use Exercise 5 above to find $\operatorname{tr}(A)$.

17. Let A be a square matrix such that $A^3 = A$. What can you say about the eigenvalues of A?

18. **a.** Solve the system

$$\begin{aligned} y_1' &= y_1 + 3y_2 \\ y_2' &= 2y_1 + 4y_2 \end{aligned}$$

b. Find the solution satisfying the initial conditions $y_1(0) = 5$ and $y_2(0) = 6$.

19. Let A be a 3×3 matrix, one of whose eigenvalues is 1. Given that both the sum and the product of all three eigenvalues is 6, what are the possible values for the remaining two eigenvalues?

20. Show that the matrices

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}$$

are similar if

$$d_k = \cos\frac{2\pi k}{3} + i\sin\frac{2\pi k}{3} \quad (k = 1, 2, 3)$$

CHAPTER 6

Inner Product Spaces

CHAPTER CONTENTS

Introduction

In Chapter 3 we defined the dot product of vectors in R^n, and we used that concept to define notions of length, angle, distance, and orthogonality. In this chapter we will generalize those ideas so they are applicable in any vector space, not just R^n. We will also discuss various applications of these ideas.

6.1 Inner Products

In this section we will use the most important properties of the dot product on R^n as axioms, which, if satisfied by the vectors in a vector space V, will enable us to extend the notions of length, distance, angle, and perpendicularity to general vector spaces.

General Inner Products

Most, but not all, of the concepts we will develop in this section apply to both real and complex vector spaces. We will limit the text discussion to real vector spaces and leave the comparable ideas for complex vector spaces for the exercises. Thus, it should be understood that all vector spaces in this section are real, even if not stated explicitly.

Definition 1

An ***inner product*** on a real vector space V is a function that associates a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ with each pair of vectors in V in such a way that the following axioms are satisfied for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and all scalars k.

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ **[Symmetry axiom]**
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$ **[Additivity axiom]**
3. $\langle k\mathbf{u}, \mathbf{v} \rangle = k\langle \mathbf{u}, \mathbf{v} \rangle$ **[Homogeneity axiom]**
4. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$ **[Positivity axiom]**

A real vector space with an inner product is called a ***real inner product space***.

Because the axioms for a real inner product space are based on properties of the dot product, these inner product space axioms will be satisfied automatically if we define the inner product of two vectors $\mathbf{u}$ and $\mathbf{v}$ in R^n to be

$$\boxed{\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n} \tag{1}$$

This inner product is commonly called the ***Euclidean inner product*** (or the ***standard inner product***) on R^n to distinguish it from other possible inner products that might be defined on R^n. We call R^n with the Euclidean inner product ***Euclidean n-space***.

Inner products can be used to define notions of norm and distance in a general inner product space just as we did with dot products in R^n. Recall from Formulas (11) and (19) of Section 3.2 that if $\mathbf{u}$ and $\mathbf{v}$ are vectors in Euclidean n-space, then norm and distance can be expressed in terms of the dot product as

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} \quad \text{and} \quad d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{(\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v})}$$

Motivated by these formulas, we make the following definition.

Definition 2

If V is a real inner product space, then the ***norm*** (or ***length***) of a vector $\mathbf{v}$ in V is denoted by $\|\mathbf{v}\|$ and is defined by

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

and the ***distance*** between two vectors is denoted by $d(\mathbf{u}, \mathbf{v})$ and is defined by

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

A vector of norm 1 is called a ***unit vector***.

The following theorem, whose proof is left for the exercises, shows that norms and distances in real inner product spaces have many of the properties that you might expect.

Theorem 6.1.1

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in a real inner product space V, and if k is a scalar, then:

(a) $\|\mathbf{v}\| \geq 0$ with equality if and only if $\mathbf{v} = \mathbf{0}$.

(b) $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$.

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

(d) $d(\mathbf{u}, \mathbf{v}) \geq 0$ with equality if and only if $\mathbf{u} = \mathbf{v}$.

Although the Euclidean inner product is the most important inner product on R^n, there are various applications in which it is desirable to modify it by *weighting* each term differently. More precisely, if

$$w_1, w_2, \ldots, w_n$$

are *positive* real numbers, called ***weights***, and if $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ are vectors in R^n, then it can be shown that the formula

$$\boxed{\langle \mathbf{u}, \mathbf{v} \rangle = w_1 u_1 v_1 + w_2 u_2 v_2 + \cdots + w_n u_n v_n} \tag{2}$$

defines an inner product on R^n that we call the ***weighted Euclidean inner product with weights $w_1, w_2, \ldots, w_n$***.

EXAMPLE 1 | Weighted Euclidean Inner Product

Let $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ be vectors in R^2. Verify that the weighted Euclidean inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1 v_1 + 2u_2 v_2 \tag{3}$$

satisfies the four inner product axioms.

Solution

Axiom 1: Interchanging $\mathbf{u}$ and $\mathbf{v}$ in Formula (3) does not change the sum on the right side, so $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.

Axiom 2: If $\mathbf{w} = (w_1, w_2)$, then

$$\begin{aligned}
\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle &= 3(u_1 + v_1)w_1 + 2(u_2 + v_2)w_2 \\
&= 3(u_1 w_1 + v_1 w_1) + 2(u_2 w_2 + v_2 w_2) \\
&= (3u_1 w_1 + 2u_2 w_2) + (3v_1 w_1 + 2v_2 w_2) \\
&= \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle
\end{aligned}$$

Axiom 3:
$$\begin{aligned}
\langle k\mathbf{u}, \mathbf{v} \rangle &= 3(ku_1)v_1 + 2(ku_2)v_2 \\
&= k(3u_1 v_1 + 2u_2 v_2) \\
&= k\langle \mathbf{u}, \mathbf{v} \rangle
\end{aligned}$$

Axiom 4: Observe that $\langle \mathbf{v}, \mathbf{v} \rangle = 3(v_1 v_1) + 2(v_2 v_2) = 3v_1^2 + 2v_2^2 \geq 0$ with equality if and only if $v_1 = v_2 = 0$, that is, if and only if $\mathbf{v} = \mathbf{0}$.

Note that the standard Euclidean inner product in Formula (1) is the special case of the weighted Euclidean inner product in which all the weights are 1.

In Example 1, we are using subscripted w's to denote the components of the vector $\mathbf{w}$, not the weights. The weights are the numbers 3 and 2 in Formula (3).

An Application of Weighted Euclidean Inner Products

To illustrate one way in which a weighted Euclidean inner product can arise, suppose that some physical experiment has n possible numerical outcomes

$$x_1, x_2, \ldots, x_n$$

and that a series of m repetitions of the experiment yields these values with various frequencies. Specifically, suppose that x_1 occurs f_1 times, x_2 occurs f_2 times, and so forth. Since there is a total of m repetitions of the experiment, it follows that

$$f_1 + f_2 + \cdots + f_n = m$$

Thus, the ***arithmetic average*** of the observed numerical values (denoted by $\bar{x}$) is

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_n x_n}{f_1 + f_2 + \cdots + f_n} = \frac{1}{m}(f_1 x_1 + f_2 x_2 + \cdots + f_n x_n) \tag{4}$$

If we let

$$\begin{aligned}
\mathbf{f} &= (f_1, f_2, \ldots, f_n) \\
\mathbf{x} &= (x_1, x_2, \ldots, x_n) \\
w_1 = w_2 &= \cdots = w_n = 1/m
\end{aligned}$$

then (4) can be expressed as the weighted Euclidean inner product

$$\bar{x} = \langle \mathbf{f}, \mathbf{x} \rangle = w_1 f_1 x_1 + w_2 f_2 x_2 + \cdots + w_n f_n x_n$$

EXAMPLE 2 | Calculating with a Weighted Euclidean Inner Product

It is important to keep in mind that norm and distance depend on the inner product being used. If the inner product is changed, then the norms and distances between vectors also change. For example, for the vectors $\mathbf{u} = (1, 0)$ and $\mathbf{v} = (0, 1)$ in R^2 with the Euclidean inner product we have

$$\|\mathbf{u}\| = \sqrt{1^2 + 0^2} = 1$$

and

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \|(1, -1)\| = \sqrt{1^2 + (-1)^2} = \sqrt{2}$$

but if we change to the weighted Euclidean inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + 2u_2v_2$$

we have

$$\|\mathbf{u}\| = \langle \mathbf{u}, \mathbf{u} \rangle^{1/2} = [3(1)(1) + 2(0)(0)]^{1/2} = \sqrt{3}$$

and

$$\begin{aligned} d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| &= \langle (1, -1), (1, -1) \rangle^{1/2} \\ &= [3(1)(1) + 2(-1)(-1)]^{1/2} = \sqrt{5} \end{aligned}$$

Unit Circles and Spheres in Inner Product Spaces

Definition 3

If V is an inner product space, then the set of points in V that satisfy

$$\|\mathbf{u}\| = 1$$

is called the ***unit sphere*** in V (or the ***unit circle*** in the case where $V = R^2$).

EXAMPLE 3 | Unusual Unit Circles in R^2

(a) Sketch the unit circle in an xy-coordinate system in R^2 using the Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 + u_2v_2$.

(b) Sketch the unit circle in an xy-coordinate system in R^2 using the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{9}u_1v_1 + \frac{1}{4}u_2v_2$.

Solution (a) If $\mathbf{u} = (x, y)$, then $\|\mathbf{u}\| = \langle \mathbf{u}, \mathbf{u} \rangle^{1/2} = \sqrt{x^2 + y^2}$, so the equation of the unit circle is $\sqrt{x^2 + y^2} = 1$, or on squaring both sides,

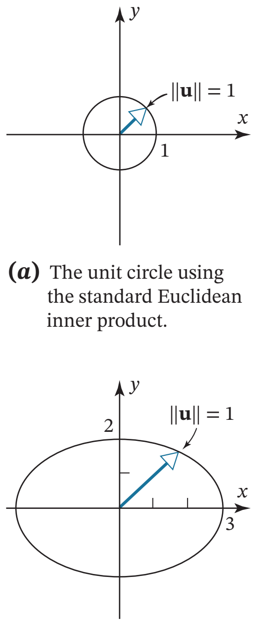
$$x^2 + y^2 = 1$$

As expected, the graph of this equation is a circle of radius 1 centered at the origin (**Figure 6.1.1a**).

Solution (b) If $\mathbf{u} = (x, y)$, then $\|\mathbf{u}\| = \langle \mathbf{u}, \mathbf{u} \rangle^{1/2} = \sqrt{\frac{1}{9}x^2 + \frac{1}{4}y^2}$, so the equation of the unit circle is $\sqrt{\frac{1}{9}x^2 + \frac{1}{4}y^2} = 1$, or on squaring both sides,

$$\frac{x^2}{9} + \frac{y^2}{4} = 1$$

The graph of this equation is the ellipse shown in Figure 6.1.1*b*. Though this may seem odd when viewed geometrically, it makes sense algebraically since all points on the ellipse are 1 unit away from the origin relative to the given weighted Euclidean inner product. In short, weighting has the effect of distorting the space that we are used to seeing through "unweighted Euclidean eyes."



(*a*) The unit circle using the standard Euclidean inner product.

(*b*) The unit circle using a weighted Euclidean inner product.

FIGURE 6.1.1

Inner Products Generated by Matrices

The Euclidean inner product and the weighted Euclidean inner products are special cases of a general class of inner products on R^n called ***matrix inner products***. To define this class of inner products, let $\mathbf{u}$ and $\mathbf{v}$ be vectors in R^n that are expressed in *column form*, and let A be an *invertible* $n \times n$ matrix. It can be shown (Exercise 47) that if $\mathbf{u} \cdot \mathbf{v}$ is the Euclidean inner product on R^n, then the formula

$$\boxed{\langle \mathbf{u}, \mathbf{v} \rangle = A\mathbf{u} \cdot A\mathbf{v}} \tag{5}$$

also defines an inner product; it is called the ***inner product on R^n generated by A***.

Recall from Table 1 of Section 3.2 that if $\mathbf{u}$ and $\mathbf{v}$ are in column form, then $\mathbf{u} \cdot \mathbf{v}$ can be written as $\mathbf{v}^T\mathbf{u}$ from which it follows that (5) can be expressed as

$$\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{v})^T A\mathbf{u}$$

or equivalently as

$$\boxed{\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^T A^T A\mathbf{u}} \tag{6}$$

EXAMPLE 4 | Matrices Generating Weighted Euclidean Inner Products

The standard Euclidean and weighted Euclidean inner products are special cases of matrix inner products. The standard Euclidean inner product on R^n is generated by the $n \times n$ identity matrix, since setting $A = I$ in Formula (5) yields

$$\langle \mathbf{u}, \mathbf{v} \rangle = I\mathbf{u} \cdot I\mathbf{v} = \mathbf{u} \cdot \mathbf{v}$$

and the weighted Euclidean inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = w_1u_1v_1 + w_2u_2v_2 + \cdots + w_nu_nv_n \tag{7}$$

is generated by the matrix

$$A = \begin{bmatrix} \sqrt{w_1} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{w_2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{w_n} \end{bmatrix}$$

This can be seen by observing that A^TA is the $n \times n$ diagonal matrix whose diagonal entries are the weights $w_1, w_2, \ldots, w_n$.

EXAMPLE 5 | Example 1 Revisited

The weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + 2u_2v_2$ discussed in Example 1 is the inner product on R^2 generated by

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

Every diagonal matrix with positive diagonal entries generates a weighted inner product. Why?

Other Examples of Inner Products

So far, we have considered only examples of inner products on R^n. We will now consider examples of inner products on some of the other kinds of vector spaces that we discussed earlier.

EXAMPLE 6 | The Standard Inner Product on M_{nn}

If $\mathbf{u} = U$ and $\mathbf{v} = V$ are matrices in the vector space M_{nn}, then the formula

$$\boxed{\langle \mathbf{u}, \mathbf{v} \rangle = \operatorname{tr}(U^T V)} \tag{8}$$

defines an inner product on M_{nn} called the ***standard inner product*** on that space (see Definition 8 of Section 1.3 for a definition of trace). This can be proved by confirming that the four inner product space axioms are satisfied, but we can illustrate the idea by computing (8) for the 2×2 matrices

$$U = \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix}$$

This yields

$$\langle \mathbf{u}, \mathbf{v} \rangle = \operatorname{tr}(U^T V) = u_1v_1 + u_2v_2 + u_3v_3 + u_4v_4$$

which is just the dot product of the corresponding entries in the two matrices. And it follows from this that

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} = \sqrt{\operatorname{tr}\langle U^T U \rangle} = \sqrt{u_1^2 + u_2^2 + u_3^2 + u_4^2}$$

For example, if

$$\mathbf{u} = U = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = V = \begin{bmatrix} -1 & 0 \\ 3 & 2 \end{bmatrix}$$

then

$$\langle \mathbf{u}, \mathbf{v} \rangle = \operatorname{tr}(U^T V) = 1(-1) + 2(0) + 3(3) + 4(2) = 16$$

and

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} = \sqrt{\operatorname{tr}(U^T U)} = \sqrt{1^2 + 2^2 + 3^2 + 4^2} = \sqrt{30}$$
$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\operatorname{tr}(V^T V)} = \sqrt{(-1)^2 + 0^2 + 3^2 + 2^2} = \sqrt{14}$$

EXAMPLE 7 | The Standard Inner Product on P_n

If

$$\mathbf{p} = a_0 + a_1x + \cdots + a_nx^n \quad \text{and} \quad \mathbf{q} = b_0 + b_1x + \cdots + b_nx^n$$

are polynomials in P_n, then the following formula defines an inner product on P_n (verify) that we will call the ***standard inner product*** on this space:

$$\langle \mathbf{p}, \mathbf{q} \rangle = a_0b_0 + a_1b_1 + \cdots + a_nb_n \tag{9}$$

The norm of a polynomial $\mathbf{p}$ relative to this inner product is

$$\|\mathbf{p}\| = \sqrt{\langle \mathbf{p}, \mathbf{p} \rangle} = \sqrt{a_0^2 + a_1^2 + \cdots + a_n^2}$$

EXAMPLE 8 | The Evaluation Inner Product on P_n

If

$$\mathbf{p} = p(x) = a_0 + a_1x + \cdots + a_nx^n \quad \text{and} \quad \mathbf{q} = q(x) = b_0 + b_1x + \cdots + b_nx^n$$

are polynomials in P_n, and if $x_0, x_1, \ldots, x_n$ are distinct real numbers (called ***sample points***), then the formula

$$\langle \mathbf{p}, \mathbf{q} \rangle = p(x_0)q(x_0) + p(x_1)q(x_1) + \cdots + p(x_n)q(x_n) \tag{10}$$

defines an inner product on P_n called the ***evaluation inner product*** at $x_0, x_1, \ldots, x_n$. Algebraically, this can be viewed as the dot product in R^n of the n-tuples

$$(p(x_0), p(x_1), \ldots, p(x_n)) \quad \text{and} \quad (q(x_0), q(x_1), \ldots, q(x_n))$$

and hence the first three inner product axioms follow from properties of the dot product. The fourth inner product axiom follows from the fact that

$$\langle \mathbf{p}, \mathbf{p} \rangle = [p(x_0)]^2 + [p(x_1)]^2 + \cdots + [p(x_n)]^2 \geq 0$$

with equality holding if and only if

$$p(x_0) = p(x_1) = \cdots = p(x_n) = 0$$

But a nonzero polynomial of degree n or less can have at most n distinct roots, so it must be that $\mathbf{p} = \mathbf{0}$, which proves that the fourth inner product axiom holds.

The norm of a polynomial $\mathbf{p}$ relative to the evaluation inner product is

$$\|\mathbf{p}\| = \sqrt{\langle \mathbf{p}, \mathbf{p} \rangle} = \sqrt{[p(x_0)]^2 + [p(x_1)]^2 + \cdots + [p(x_n)]^2} \tag{11}$$

EXAMPLE 9 | Working with the Evaluation Inner Product

Let P_2 have the evaluation inner product at the points

$$x_0 = -2, \quad x_1 = 0, \quad \text{and} \quad x_2 = 2$$

Compute $\langle \mathbf{p}, \mathbf{q} \rangle$ and $\|\mathbf{p}\|$ for the polynomials $\mathbf{p} = p(x) = x^2$ and $\mathbf{q} = q(x) = 1 + x$.

Solution It follows from (10) and (11) that

$$\langle \mathbf{p}, \mathbf{q} \rangle = p(-2)q(-2) + p(0)q(0) + p(2)q(2) = (4)(-1) + (0)(1) + (4)(3) = 8$$

$$\begin{aligned} \|\mathbf{p}\| &= \sqrt{[p(x_0)]^2 + [p(x_1)]^2 + [p(x_2)]^2} = \sqrt{[p(-2)]^2 + [p(0)]^2 + [p(2)]^2} \\ &= \sqrt{4^2 + 0^2 + 4^2} = \sqrt{32} = 4\sqrt{2} \end{aligned}$$

EXAMPLE 10 | An Integral Inner Product on $C[a, b]$

CALCULUS REQUIRED

Let $\mathbf{f} = f(x)$ and $\mathbf{g} = g(x)$ be two functions in $C[a, b]$ and define

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x)\, dx \tag{12}$$

We will show that this formula defines an inner product on $C[a, b]$ by verifying the four inner product axioms for functions $\mathbf{f} = f(x)$, $\mathbf{g} = g(x)$, and $\mathbf{h} = h(x)$ in $C[a, b]$.

Axiom 1: $\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x)\,dx = \int_a^b g(x)f(x)\,dx = \langle \mathbf{g}, \mathbf{f} \rangle$

Axiom 2: $$\begin{aligned} \langle \mathbf{f} + \mathbf{g}, \mathbf{h} \rangle &= \int_a^b (f(x) + g(x))h(x)\,dx \\ &= \int_a^b f(x)h(x)\,dx + \int_a^b g(x)h(x)\,dx \\ &= \langle \mathbf{f}, \mathbf{h} \rangle + \langle \mathbf{g}, \mathbf{h} \rangle \end{aligned}$$

Axiom 3: $\langle k\mathbf{f}, \mathbf{g} \rangle = \int_a^b kf(x)g(x)\,dx = k\int_a^b f(x)g(x)\,dx = k\langle \mathbf{f}, \mathbf{g} \rangle$

Axiom 4: If $\mathbf{f} = f(x)$ is any function in $C[a, b]$, then

$$\langle \mathbf{f}, \mathbf{f} \rangle = \int_a^b f^2(x)\,dx \geq 0 \tag{13}$$

since $f^2(x) \geq 0$ for all x in the interval $[a, b]$. Moreover, because f is continuous on $[a, b]$, the equality in Formula (13) holds if and only if the function f is identically zero on $[a, b]$, that is, if and only if $\mathbf{f} = \mathbf{0}$; and this proves that Axiom 4 holds.

CALCULUS REQUIRED

EXAMPLE 11 | Norm of a Vector in $C[a, b]$

If $C[a, b]$ has the inner product that was defined in Example 10, then the norm of a function $\mathbf{f} = f(x)$ relative to this inner product is

$$\|\mathbf{f}\| = \langle \mathbf{f}, \mathbf{f} \rangle^{1/2} = \sqrt{\int_a^b f^2(x)\,dx} \tag{14}$$

and the unit sphere in this space consists of all functions $\mathbf{f}$ in $C[a, b]$ that satisfy the equation

$$\int_a^b f^2(x)\,dx = 1$$

Remark Note that the vector space P_n is a subspace of $C[a, b]$ because polynomials are continuous functions. Thus, Formula (12) defines an inner product on P_n that is different from both the standard inner product and the evaluation inner product.

Warning Recall from calculus that the arc length of a curve $y = f(x)$ over an interval $[a, b]$ is given by the formula

$$L = \int_a^b \sqrt{1 + [f'(x)]^2}\,dx \tag{15}$$

Do not confuse this concept of arc length with $\|\mathbf{f}\|$, which is the length (norm) of $\mathbf{f}$ when $\mathbf{f}$ is viewed as a vector in $C[a, b]$. Formulas (14) and (15) have different meanings.

Algebraic Properties of Inner Products

The following theorem lists some of the algebraic properties of inner products that follow from the inner product axioms. This result is a generalization of Theorem 3.2.3, which applied only to the dot product on R^n.

Theorem 6.1.2

If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in a real inner product space V, and if k is a scalar, then:

(a) $\langle \mathbf{0}, \mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{0}\rangle = 0$

(b) $\langle \mathbf{u}, \mathbf{v}+\mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{u}, \mathbf{w}\rangle$

(c) $\langle \mathbf{u}, \mathbf{v}-\mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle - \langle \mathbf{u}, \mathbf{w}\rangle$

(d) $\langle \mathbf{u}-\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{w}\rangle - \langle \mathbf{v}, \mathbf{w}\rangle$

(e) $k\langle \mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, k\mathbf{v}\rangle$

Proof We will prove part (b) and leave the proofs of the remaining parts for the reader.

$$\begin{aligned}
\langle \mathbf{u}, \mathbf{v}+\mathbf{w}\rangle &= \langle \mathbf{v}+\mathbf{w}, \mathbf{u}\rangle && \textbf{[By symmetry]}\\
&= \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{w}, \mathbf{u}\rangle && \textbf{[By additivity]}\\
&= \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{u}, \mathbf{w}\rangle && \textbf{[By symmetry]}
\end{aligned}$$

The following example illustrates how Theorem 6.1.2 and the defining properties of inner products can be used to perform algebraic computations with inner products. As you read through the example, you will find it instructive to justify the steps.

EXAMPLE 12 | Calculating with Inner Products

$$\begin{aligned}
\langle \mathbf{u}-2\mathbf{v}, 3\mathbf{u}+4\mathbf{v}\rangle &= \langle \mathbf{u}, 3\mathbf{u}+4\mathbf{v}\rangle - \langle 2\mathbf{v}, 3\mathbf{u}+4\mathbf{v}\rangle\\
&= \langle \mathbf{u}, 3\mathbf{u}\rangle + \langle \mathbf{u}, 4\mathbf{v}\rangle - \langle 2\mathbf{v}, 3\mathbf{u}\rangle - \langle 2\mathbf{v}, 4\mathbf{v}\rangle\\
&= 3\langle \mathbf{u}, \mathbf{u}\rangle + 4\langle \mathbf{u}, \mathbf{v}\rangle - 6\langle \mathbf{v}, \mathbf{u}\rangle - 8\langle \mathbf{v}, \mathbf{v}\rangle\\
&= 3\|\mathbf{u}\|^2 + 4\langle \mathbf{u}, \mathbf{v}\rangle - 6\langle \mathbf{u}, \mathbf{v}\rangle - 8\|\mathbf{v}\|^2\\
&= 3\|\mathbf{u}\|^2 - 2\langle \mathbf{u}, \mathbf{v}\rangle - 8\|\mathbf{v}\|^2
\end{aligned}$$

Exercise Set 6.1

1. Let R^2 have the weighted Euclidean inner product

$$\langle \mathbf{u}, \mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2$$

and let $\mathbf{u} = (1, 1)$, $\mathbf{v} = (3, 2)$, $\mathbf{w} = (0, -1)$, and $k = 3$. Compute the stated quantities.

a. $\langle \mathbf{u}, \mathbf{v}\rangle$ **b.** $\langle k\mathbf{v}, \mathbf{w}\rangle$ **c.** $\langle \mathbf{u}+\mathbf{v}, \mathbf{w}\rangle$

d. $\|\mathbf{v}\|$ **e.** $d(\mathbf{u}, \mathbf{v})$ **f.** $\|\mathbf{u} - k\mathbf{v}\|$

2. Follow the directions of Exercise 1 using the weighted Euclidean inner product

$$\langle \mathbf{u}, \mathbf{v}\rangle = \tfrac{1}{2}u_1v_1 + 5u_2v_2$$

In Exercises 3–4, compute the quantities in parts (a)–(f) *of Exercise 1 using the inner product on R^2 generated by A.*

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

In Exercises 5–6, find a matrix that generates the stated weighted inner product on R^2.

5. $\langle \mathbf{u}, \mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2$ **6.** $\langle \mathbf{u}, \mathbf{v}\rangle = \frac{1}{2}u_1v_1 + 5u_2v_2$

In Exercises 7–8, use the inner product on R^2 generated by the matrix A to find $\langle \mathbf{u}, \mathbf{v}\rangle$ for the vectors $\mathbf{u} = (0, -3)$ and $\mathbf{v} = (6, 2)$.

7. $A = \begin{bmatrix} 4 & 1 \\ 2 & -3 \end{bmatrix}$ **8.** $A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}$

In Exercises 9–10, compute the standard inner product on M_{22} of the given matrices.

9. $U = \begin{bmatrix} 3 & -2 \\ 4 & 8 \end{bmatrix}$, $V = \begin{bmatrix} -1 & 3 \\ 1 & 1 \end{bmatrix}$

10. $U = \begin{bmatrix} 1 & 2 \\ -3 & 5 \end{bmatrix}, \quad V = \begin{bmatrix} 4 & 6 \\ 0 & 8 \end{bmatrix}$

In Exercises 11–12, find the standard inner product on P_2 of the given polynomials.

11. $\mathbf{p} = -2 + x + 3x^2, \quad \mathbf{q} = 4 - 7x^2$

12. $\mathbf{p} = -5 + 2x + x^2, \quad \mathbf{q} = 3 + 2x - 4x^2$

In Exercises 13–14, a weighted Euclidean inner product on R^2 is given for the vectors $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$. Find a matrix that generates it.

13. $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + 5u_2v_2$ **14.** $\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + 6u_2v_2$

In Exercises 15–16, a sequence of sample points is given. Use the evaluation inner product on P_3 at those sample points to find $\langle \mathbf{p}, \mathbf{q} \rangle$ for the polynomials

$$\mathbf{p} = x + x^3 \quad \textit{and} \quad \mathbf{q} = 1 + x^2$$

15. $x_0 = -2, \; x_1 = -1, \; x_2 = 0, \; x_3 = 1$

16. $x_0 = -1, \; x_1 = 0, \; x_2 = 1, \; x_3 = 2$

In Exercises 17–18, find $\|\mathbf{u}\|$ and $d(\mathbf{u}, \mathbf{v})$ relative to the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + 3u_2v_2$ on R^2.

17. $\mathbf{u} = (-3, 2)$ and $\mathbf{v} = (1, 7)$

18. $\mathbf{u} = (-1, 2)$ and $\mathbf{v} = (2, 5)$

In Exercises 19–20, find $\|\mathbf{p}\|$ and $d(\mathbf{p}, \mathbf{q})$ relative to the standard inner product on P_2.

19. $\mathbf{p} = -2 + x + 3x^2, \quad \mathbf{q} = 4 - 7x^2$

20. $\mathbf{p} = -5 + 2x + x^2, \quad \mathbf{q} = 3 + 2x - 4x^2$

In Exercises 21–22, find $\|U\|$ and $d(U, V)$ relative to the standard inner product on M_{22}.

21. $U = \begin{bmatrix} 3 & -2 \\ 4 & 8 \end{bmatrix}, \quad V = \begin{bmatrix} -1 & 3 \\ 1 & 1 \end{bmatrix}$

22. $U = \begin{bmatrix} 1 & 2 \\ -3 & 5 \end{bmatrix}, \quad V = \begin{bmatrix} 4 & 6 \\ 0 & 8 \end{bmatrix}$

In Exercises 23–24, let

$$\mathbf{p} = x + x^3 \quad \textit{and} \quad \mathbf{q} = 1 + x^2$$

Find $\|\mathbf{p}\|$ and $d(\mathbf{p}, \mathbf{q})$ relative to the evaluation inner product on P_3 at the stated sample points.

23. $x_0 = -2, \; x_1 = -1, \; x_2 = 0, \; x_3 = 1$

24. $x_0 = -1, \; x_1 = 0, \; x_2 = 1, \; x_3 = 2$

In Exercises 25–26, find $\|\mathbf{u}\|$ and $d(\mathbf{u}, \mathbf{v})$ for the vectors $\mathbf{u} = (-1, 2)$ and $\mathbf{v} = (2, 5)$ relative to the inner product on R^2 generated by the matrix A.

25. $A = \begin{bmatrix} 4 & 0 \\ 3 & 5 \end{bmatrix}$ **26.** $A = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}$

In Exercises 27–28, suppose that $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in an inner product space such that

$$\langle \mathbf{u}, \mathbf{v} \rangle = 2, \quad \langle \mathbf{v}, \mathbf{w} \rangle = -6, \quad \langle \mathbf{u}, \mathbf{w} \rangle = -3$$
$$\|\mathbf{u}\| = 1, \quad \|\mathbf{v}\| = 2, \quad \|\mathbf{w}\| = 7$$

Evaluate the given expression.

27. a. $\langle 2\mathbf{v} - \mathbf{w}, 3\mathbf{u} + 2\mathbf{w} \rangle$ **b.** $\|\mathbf{u} + \mathbf{v}\|$

28. a. $\langle \mathbf{u} - \mathbf{v} - 2\mathbf{w}, 4\mathbf{u} + \mathbf{v} \rangle$ **b.** $\|2\mathbf{w} - \mathbf{v}\|$

In Exercises 29–30, sketch the unit circle in R^2 using the given inner product.

29. $\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{4}u_1v_1 + \frac{1}{16}u_2v_2$ **30.** $\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + u_2v_2$

In Exercises 31–32, find a weighted Euclidean inner product on R^2 for which the "unit circle" is the ellipse shown in the accompanying figure.

31.

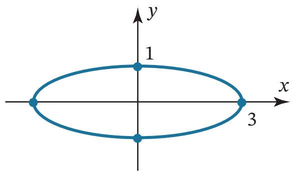


FIGURE Ex-31

32.

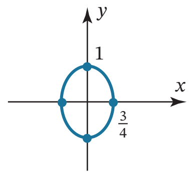


FIGURE Ex-32

In Exercises 33–34, let $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$. Show that the expression does not *define an inner product on R^3, and list all inner product axioms that fail to hold.*

33. $\langle \mathbf{u}, \mathbf{v} \rangle = u_1^2v_1^2 + u_2^2v_2^2 + u_3^2v_3^2$

34. $\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 - u_2v_2 + u_3v_3$

In Exercises 35–36, suppose that $\mathbf{u}$ and $\mathbf{v}$ are vectors in an inner product space. Rewrite the given expression in terms of $\langle \mathbf{u}, \mathbf{v} \rangle$, $\|\mathbf{u}\|^2$, and $\|\mathbf{v}\|^2$.

35. $\langle 2\mathbf{v} - 4\mathbf{u}, \mathbf{u} - 3\mathbf{v} \rangle$ **36.** $\langle 5\mathbf{u} + 6\mathbf{v}, 4\mathbf{v} - 3\mathbf{u} \rangle$

37. **(*Calculus required*)** Let the vector space P_2 have the inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_{-1}^{1} p(x)q(x)\,dx$$

Find the following for $\mathbf{p} = 1$ and $\mathbf{q} = x^2$.

a. $\langle \mathbf{p}, \mathbf{q} \rangle$ **b.** $d(\mathbf{p}, \mathbf{q})$

c. $\|\mathbf{p}\|$ **d.** $\|\mathbf{q}\|$

38. **(*Calculus required*)** Let the vector space P_3 have the inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_{-1}^{1} p(x)q(x)\,dx$$

Find the following for $\mathbf{p} = 2x^3$ and $\mathbf{q} = 1 - x^3$.

a. $\langle \mathbf{p}, \mathbf{q} \rangle$ **b.** $d(\mathbf{p}, \mathbf{q})$

c. $\|\mathbf{p}\|$ **d.** $\|\mathbf{q}\|$

(*Calculus required*) *In Exercises 39–40, use the inner product*

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_{0}^{1} f(x)g(x)dx$$

on $C[0, 1]$ to compute $\langle \mathbf{f}, \mathbf{g} \rangle$.

39. $\mathbf{f} = \cos 2\pi x, \quad \mathbf{g} = \sin 2\pi x$ **40.** $\mathbf{f} = x, \quad \mathbf{g} = e^x$

Working with Proofs

41. Prove parts (a) and (b) of Theorem 6.1.1.

42. Prove parts (c) and (d) of Theorem 6.1.1.

43. a. Let $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$. Prove that the expression $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + 5u_2v_2$ defines an inner product on R^2 by showing that the inner product axioms hold.

b. What conditions must k_1 and k_2 satisfy for the expression $\langle \mathbf{u}, \mathbf{v} \rangle = k_1u_1v_1 + k_2u_2v_2$ to define an inner product on R^2? Justify your answer.

44. Prove that the following identity holds for vectors in any inner product space.

$$\langle \mathbf{u}, \mathbf{v} \rangle = \tfrac{1}{4}\|\mathbf{u} + \mathbf{v}\|^2 - \tfrac{1}{4}\|\mathbf{u} - \mathbf{v}\|^2$$

45. Prove that the following identity holds for vectors in any inner product space.

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

46. The definition of a complex vector space was given in the first margin note in Section 4.1. The definition of a ***complex inner product*** on a complex vector space V is identical to that in Definition 1 except that scalars are allowed to be complex numbers, and Axiom 1 is replaced by $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$. The remaining axioms are unchanged. A complex vector space with a complex inner product is called a ***complex inner product space***. Prove that if V is a complex inner product space, then $\langle \mathbf{u}, k\mathbf{v} \rangle = \overline{k}\langle \mathbf{u}, \mathbf{v} \rangle$.

47. Prove that Formula (5) defines an inner product on R^n.

48. a. Prove that if $\mathbf{v}$ is a fixed vector in a real inner product space V, then the mapping $T: V \to R$ defined by $T(\mathbf{x}) = \langle \mathbf{x}, \mathbf{v} \rangle$ is a linear transformation.

b. Let $V = R^3$ have the Euclidean inner product, and let $\mathbf{v}$ be the vector $(1, 0, 2)$. Compute $T(1, 1, 1)$.

c. Let $V = P_2$ have the standard inner product, and let $\mathbf{v}$ be the vector $1 + x$. Compute $T(x + x^2)$.

d. Let $V = P_2$ have the evaluation inner product at the points $x_0 = 1$, $x_1 = 0$, $x_2 = -1$, and let $\mathbf{v} = 1 + x$. Compute $T(x + x^2)$.

True-False Exercises

TF. In parts **(a)–(g)** determine whether the statement is true or false, and justify your answer.

a. The dot product on R^2 is an example of a weighted inner product.

b. The inner product of two vectors cannot be a negative real number.

c. $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{u} \rangle$.

d. $\langle k\mathbf{u}, k\mathbf{v} \rangle = k^2\langle \mathbf{u}, \mathbf{v} \rangle$.

e. If $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, then $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$.

f. If $\|\mathbf{v}\|^2 = 0$, then $\mathbf{v} = \mathbf{0}$.

g. If A is an $n \times n$ matrix, then $\langle \mathbf{u}, \mathbf{v} \rangle = A\mathbf{u} \cdot A\mathbf{v}$ defines an inner product on R^n.

Working with Technology

T1. a. Confirm that the following matrix generates an inner product.

$$A = \begin{bmatrix} 5 & 8 & 6 & -13 \\ 3 & -1 & 0 & -9 \\ 0 & 1 & -1 & 0 \\ 2 & 4 & 3 & -5 \end{bmatrix}$$

b. For the following vectors, use the inner product in part (a) to compute $\langle \mathbf{u}, \mathbf{v} \rangle$, first by Formula (5) and then by Formula (6).

$$\mathbf{u} = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 3 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 2 \end{bmatrix}$$

T2. Let the vector space P_4 have the evaluation inner product at the points

$$-2, \quad -1, \quad 0, \quad 1, \quad 2$$

and let

$$\mathbf{p} = p(x) = x + x^3 \quad \text{and} \quad \mathbf{q} = q(x) = 1 + x^2 + x^4$$

a. Compute $\langle \mathbf{p}, \mathbf{q} \rangle$, $\|\mathbf{p}\|$, and $\|\mathbf{q}\|$.

b. Verify that the identities in Exercises 44 and 45 hold for the vectors $\mathbf{p}$ and $\mathbf{q}$.

T3. Let the vector space M_{33} have the standard inner product and let

$$\mathbf{u} = U = \begin{bmatrix} 1 & -2 & 3 \\ -2 & 4 & 1 \\ 3 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = V = \begin{bmatrix} 2 & -1 & 0 \\ 1 & 4 & 3 \\ 1 & 0 & 2 \end{bmatrix}$$

a. Use Formula (8) to compute $\langle \mathbf{u}, \mathbf{v} \rangle$, $\|\mathbf{u}\|$, and $\|\mathbf{v}\|$.

b. Verify that the identities in Exercises 44 and 45 hold for the vectors $\mathbf{u}$ and $\mathbf{v}$.

6.2 Angle and Orthogonality in Inner Product Spaces

In Section 3.2 we defined the notion of "angle" between vectors in R^n. In this section we will extend this idea to general vector spaces. This will enable us to extend the notion of orthogonality as well, thereby setting the groundwork for a variety of new applications.

Cauchy–Schwarz Inequality

Recall from Formula (20) of Section 3.2 that the angle θ between two vectors $\mathbf{u}$ and $\mathbf{v}$ in R^n is

$$\theta = \cos^{-1}\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}\right) \tag{1}$$

We were assured that this formula was valid because it followed from the Cauchy–Schwarz inequality (Theorem 3.2.4) that

$$-1 \le \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} \le 1 \tag{2}$$

as required for the inverse cosine to be defined. The following generalization of the Cauchy–Schwarz inequality will enable us to define the angle between two vectors in *any* real inner product space.

Theorem 6.2.1

Cauchy–Schwarz Inequality

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in a real inner product space V, then

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \le \|\mathbf{u}\|\|\mathbf{v}\| \tag{3}$$

Proof We warn you in advance that the proof presented here depends on a clever trick that is not easy to motivate.

In the case where $\mathbf{u} = \mathbf{0}$ the two sides of (3) are equal since $\langle \mathbf{u}, \mathbf{v}\rangle$ and $\|\mathbf{u}\|$ are both zero. Thus, we need consider only the case where $\mathbf{u} \neq \mathbf{0}$. Making this assumption, let

$$a = \langle \mathbf{u}, \mathbf{u}\rangle, \quad b = 2\langle \mathbf{u}, \mathbf{v}\rangle, \quad c = \langle \mathbf{v}, \mathbf{v}\rangle$$

and let t be any real number. Since the positivity axiom states that the inner product of any vector with itself is nonnegative, it follows that

$$0 \le \langle t\mathbf{u} + \mathbf{v}, t\mathbf{u} + \mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{u}\rangle t^2 + 2\langle \mathbf{u}, \mathbf{v}\rangle t + \langle \mathbf{v}, \mathbf{v}\rangle = at^2 + bt + c$$

This inequality implies that the quadratic polynomial $at^2 + bt + c$ has either no real roots or a repeated real root. Therefore, its discriminant must satisfy the inequality $b^2 - 4ac \le 0$. Expressing the coefficients a, b, and c in terms of the vectors $\mathbf{u}$ and $\mathbf{v}$ gives

$$4\langle \mathbf{u}, \mathbf{v}\rangle^2 - 4\langle \mathbf{u}, \mathbf{u}\rangle\langle \mathbf{v}, \mathbf{v}\rangle \le 0$$

or, equivalently,

$$\langle \mathbf{u}, \mathbf{v}\rangle^2 \le \langle \mathbf{u}, \mathbf{u}\rangle\langle \mathbf{v}, \mathbf{v}\rangle$$

Taking square roots of both sides and using the fact that $\langle \mathbf{u}, \mathbf{u}\rangle$ and $\langle \mathbf{v}, \mathbf{v}\rangle$ are nonnegative yields

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \le \langle \mathbf{u}, \mathbf{u}\rangle^{1/2}\langle \mathbf{v}, \mathbf{v}\rangle^{1/2} \quad \text{or equivalently} \quad |\langle \mathbf{u}, \mathbf{v}\rangle| \le \|\mathbf{u}\|\|\mathbf{v}\|$$

which completes the proof. ■

The following two alternative forms of the Cauchy–Schwarz inequality are useful to know:

$$\boxed{\langle \mathbf{u}, \mathbf{v}\rangle^2 \leq \langle \mathbf{u}, \mathbf{u}\rangle\langle \mathbf{v}, \mathbf{v}\rangle} \tag{4}$$

$$\boxed{\langle \mathbf{u}, \mathbf{v}\rangle^2 \leq \|\mathbf{u}\|^2\|\mathbf{v}\|^2} \tag{5}$$

The first of these formulas was obtained in the proof of Theorem 6.2.1, and the second is a variation of the first.

Angle Between Vectors

Our next goal is to define what is meant by the "angle" between vectors in a real inner product space. As a first step, we leave it as an exercise for you to use the Cauchy–Schwarz inequality to show that

$$-1 \leq \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} \leq 1 \tag{6}$$

This being the case, there is a unique angle θ in radian measure for which

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} \quad \text{and} \quad 0 \leq \theta \leq \pi \tag{7}$$

(**Figure 6.2.1**). This enables us to *define* the ***angle θ between u and v*** to be

$$\boxed{\theta = \cos^{-1}\left(\frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}\right)} \tag{8}$$

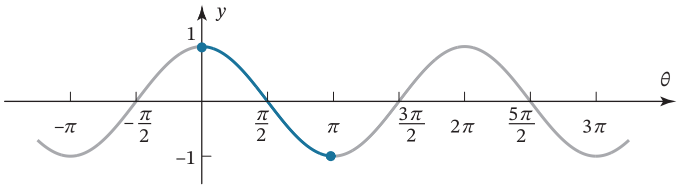


FIGURE 6.2.1

EXAMPLE 1 | Cosine of the Angle Between Vectors in M_{22}

Let M_{22} have the standard inner product. Find the cosine of the angle between the vectors

$$\mathbf{u} = U = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = V = \begin{bmatrix} -1 & 0 \\ 3 & 2 \end{bmatrix}$$

Solution We showed in Example 6 of the previous section that

$$\langle \mathbf{u}, \mathbf{v}\rangle = 16, \quad \|\mathbf{u}\| = \sqrt{30}, \quad \|\mathbf{v}\| = \sqrt{14}$$

from which it follows that

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{16}{\sqrt{30}\sqrt{14}} \approx 0.78$$

Properties of Length and Distance in General Inner Product Spaces

In Section 3.2 we used the dot product to extend the notions of length and distance to R^n, and we showed that various basic geometry theorems remained valid (see Theorems 3.2.5, 3.2.6, and 3.2.7). By making only minor adjustments to the proofs of those theorems, one can show that they remain valid in any real inner product space. For example, here is the generalization of Theorem 3.2.5 (the triangle inequalities).

Theorem 6.2.2

If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in a real inner product space V, and if k is any scalar, then:

(*a*) $\|\mathbf{u}+\mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|$ **[Triangle inequality for vectors]**

(*b*) $d(\mathbf{u},\mathbf{v}) \le d(\mathbf{u},\mathbf{w}) + d(\mathbf{w},\mathbf{v})$ **[Triangle inequality for distances]**

Proof (a)

$$\begin{aligned}
\|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle \\
&= \langle \mathbf{u},\mathbf{u}\rangle + 2\langle \mathbf{u},\mathbf{v}\rangle + \langle \mathbf{v},\mathbf{v}\rangle \\
&\le \langle \mathbf{u},\mathbf{u}\rangle + 2|\langle \mathbf{u},\mathbf{v}\rangle| + \langle \mathbf{v},\mathbf{v}\rangle \quad \textbf{[Property of absolute value]} \\
&\le \langle \mathbf{u},\mathbf{u}\rangle + 2\|\mathbf{u}\|\|\mathbf{v}\| + \langle \mathbf{v},\mathbf{v}\rangle \quad \textbf{[By (3)]} \\
&= \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\
&= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2
\end{aligned}$$

Taking square roots gives $\|\mathbf{u}+\mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|$.

Proof (b) Identical to the proof of part (*b*) of Theorem 3.2.5. ■

Orthogonality

Although Example 1 is a useful mathematical exercise, there is only an occasional need to compute angles in vector spaces other than R^2 and R^3. A problem of more importance in general vector spaces is ascertaining whether the angle between vectors is $\pi/2$. You should be able to see from Formula (8) that if $\mathbf{u}$ and $\mathbf{v}$ are *nonzero* vectors, then the angle between them is $\theta = \pi/2$ if and only if $\langle \mathbf{u},\mathbf{v}\rangle = 0$. Accordingly, we make the following definition, which is a generalization of Definition 1 in Section 3.3 and is applicable even if one or both of the vectors is zero.

Definition 1

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in an inner product space V are called ***orthogonal*** if $\langle \mathbf{u},\mathbf{v}\rangle = 0$.

As the following example shows, orthogonality depends on the inner product in the sense that for different inner products two vectors can be orthogonal with respect to one but not the other.

EXAMPLE 2 | Orthogonality Depends on the Inner Product

The vectors $\mathbf{u} = (1, 1)$ and $\mathbf{v} = (1, -1)$ are orthogonal with respect to the Euclidean inner product on R^2 since

$$\mathbf{u} \cdot \mathbf{v} = (1)(1) + (1)(-1) = 0$$

However, they are not orthogonal with respect to the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + 2u_2v_2$ since

$$\langle \mathbf{u}, \mathbf{v} \rangle = 3(1)(1) + 2(1)(-1) = 1 \neq 0$$

EXAMPLE 3 | Orthogonal Vectors in M_{22}

If M_{22} has the inner product of Example 6 in the preceding section, then the matrices

$$U = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$$

are orthogonal when viewed as vectors since

$$\langle U, V \rangle = 1(0) + 0(2) + 1(0) + 1(0) = 0$$

EXAMPLE 4 | Orthogonal Vectors in P_2

CALCULUS REQUIRED

Let P_2 have the inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_{-1}^{1} p(x)q(x)\, dx$$

and let $\mathbf{p} = x$ and $\mathbf{q} = x^2$. Then

$$\|\mathbf{p}\| = \langle \mathbf{p}, \mathbf{p} \rangle^{1/2} = \left[\int_{-1}^{1} xx\, dx\right]^{1/2} = \left[\int_{-1}^{1} x^2\, dx\right]^{1/2} = \sqrt{\frac{2}{3}}$$

$$\|\mathbf{q}\| = \langle \mathbf{q}, \mathbf{q} \rangle^{1/2} = \left[\int_{-1}^{1} x^2x^2\, dx\right]^{1/2} = \left[\int_{-1}^{1} x^4\, dx\right]^{1/2} = \sqrt{\frac{2}{5}}$$

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_{-1}^{1} xx^2\, dx = \int_{-1}^{1} x^3\, dx = 0$$

Because $\langle \mathbf{p}, \mathbf{q} \rangle = 0$, the vectors $\mathbf{p} = x$ and $\mathbf{q} = x^2$ are orthogonal relative to the given integral inner product.

In Theorem 3.3.3 we proved the Theorem of Pythagoras for vectors in Euclidean n-space. The following theorem extends this result to vectors in any real inner product space.

Theorem 6.2.3

Generalized Theorem of Pythagoras

If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal vectors in a real inner product space, then

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

Proof The orthogonality of $\mathbf{u}$ and $\mathbf{v}$ implies that $\langle \mathbf{u}, \mathbf{v}\rangle = 0$, so

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle = \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v}\rangle + \|\mathbf{v}\|^2 \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \quad \blacksquare\end{aligned}$$

CALCULUS REQUIRED

EXAMPLE 5 | Theorem of Pythagoras in P_2

In Example 4 we showed that $\mathbf{p} = x$ and $\mathbf{q} = x^2$ are orthogonal with respect to the inner product

$$\langle \mathbf{p}, \mathbf{q}\rangle = \int_{-1}^{1} p(x)q(x)\,dx$$

on P_2. It follows from Theorem 6.2.3 that

$$\|\mathbf{p}+\mathbf{q}\|^2 = \|\mathbf{p}\|^2 + \|\mathbf{q}\|^2$$

Thus, from the computations in Example 4, we have

$$\|\mathbf{p}+\mathbf{q}\|^2 = \left(\sqrt{\frac{2}{3}}\right)^2 + \left(\sqrt{\frac{2}{5}}\right)^2 = \frac{2}{3} + \frac{2}{5} = \frac{16}{15}$$

We can check this result by direct integration:

$$\begin{aligned}\|\mathbf{p}+\mathbf{q}\|^2 = \langle \mathbf{p}+\mathbf{q}, \mathbf{p}+\mathbf{q}\rangle &= \int_{-1}^{1} (x+x^2)(x+x^2)\,dx \\ &= \int_{-1}^{1} x^2\,dx + 2\int_{-1}^{1} x^3\,dx + \int_{-1}^{1} x^4\,dx = \frac{2}{3} + 0 + \frac{2}{5} = \frac{16}{15}\end{aligned}$$

Orthogonal Complements

In Section 4.9 we defined the notion of an *orthogonal complement* for subspaces of R^n, and we used that definition to establish a geometric link between the fundamental spaces of a matrix. The following definition extends that idea to general inner product spaces.

Definition 2

If W is a subspace of a real inner product space V, then the set of all vectors in V that are orthogonal to every vector in W is called the ***orthogonal complement*** of W and is denoted by the symbol $W^{\perp}$.

In Theorem 4.9.6 we stated three properties of orthogonal complements in R^n. The following theorem generalizes parts (a) and (b) of that theorem to general real inner product spaces.

Theorem 6.2.4

If W is a subspace of a real inner product space V, then:

(*a*) $W^\perp$ is a subspace of V.

(*b*) $W \cap W^\perp = \{\mathbf{0}\}$.

Proof (a) The set $W^\perp$ contains at least the zero vector, since $\langle \mathbf{0}, \mathbf{w} \rangle = 0$ for every vector $\mathbf{w}$ in W. Thus, it remains to show that $W^\perp$ is closed under addition and scalar multiplication. To do this, suppose that $\mathbf{u}$ and $\mathbf{v}$ are vectors in $W^\perp$, so that for every vector $\mathbf{w}$ in W we have $\langle \mathbf{u}, \mathbf{w} \rangle = 0$ and $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. It follows from the additivity and homogeneity axioms of inner products that

$$\langle \mathbf{u}+\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle = 0 + 0 = 0$$
$$\langle k\mathbf{u}, \mathbf{w} \rangle = k\langle \mathbf{u}, \mathbf{w} \rangle = k(0) = 0$$

which proves that $\mathbf{u}+\mathbf{v}$ and $k\mathbf{u}$ are in $W^\perp$.

Proof (b) If $\mathbf{v}$ is any vector in both W and $W^\perp$, then $\mathbf{v}$ is orthogonal to itself; that is, $\langle \mathbf{v}, \mathbf{v} \rangle = 0$. It follows from the positivity axiom for inner products that $\mathbf{v} = \mathbf{0}$. ■

The next theorem, which we state without proof, generalizes part (*c*) of Theorem 4.9.6. Note, however, that this theorem applies only to finite-dimensional inner product spaces, whereas Theorem 4.9.6 does not have this restriction.

Theorem 6.2.5

If W is a subspace of a real finite-dimensional inner product space V, then the orthogonal complement of $W^\perp$ is W; that is,

$$(W^\perp)^\perp = W$$

Theorem 6.2.5 implies that in a finite-dimensional inner product space orthogonal complements occur in pairs, each being orthogonal to the other (**Figure 6.2.2**).

In our study of the fundamental spaces of a matrix in Section 4.9 we showed that the row space and null space of a matrix are orthogonal complements with respect to the Euclidean inner product on R^n (Theorem 4.9.7). The following example takes advantage of that fact.

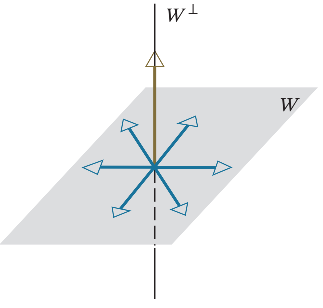


FIGURE 6.2.2 Each vector in W is orthogonal to each vector in $W^\perp$ and conversely.

EXAMPLE 6 | Basis for an Orthogonal Complement

Let W be the subspace of R^6 spanned by the vectors

$$\mathbf{w}_1 = (1, 3, -2, 0, 2, 0), \quad \mathbf{w}_2 = (2, 6, -5, -2, 4, -3),$$
$$\mathbf{w}_3 = (0, 0, 5, 10, 0, 15), \quad \mathbf{w}_4 = (2, 6, 0, 8, 4, 18)$$

Find a basis for the orthogonal complement of W.

Solution The subspace W is the same as the row space of the matrix

$$A = \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 \\ 0 & 0 & 5 & 10 & 0 & 15 \\ 2 & 6 & 0 & 8 & 4 & 18 \end{bmatrix}$$

Since the row space and null space of A are orthogonal complements, our problem reduces to finding a basis for the null space of this matrix. In Example 4 of Section 4.8 we showed that

$$\mathbf{v}_1 = \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -4 \\ 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

form a basis for this null space. Expressing these vectors in comma-delimited form (to match that of $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$, and $\mathbf{w}_4$), we obtain the basis vectors

$$\mathbf{v}_1 = (-3, 1, 0, 0, 0, 0), \quad \mathbf{v}_2 = (-4, 0, -2, 1, 0, 0), \quad \mathbf{v}_3 = (-2, 0, 0, 0, 1, 0)$$

You may want to check that these vectors are orthogonal to $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$, and $\mathbf{w}_4$ by computing the necessary dot products.

Exercise Set 6.2

In Exercises 1–2, find the cosine of the angle between the vectors with respect to the Euclidean inner product.

1. **a.** $\mathbf{u} = (1, -3)$, $\mathbf{v} = (2, 4)$

b. $\mathbf{u} = (-1, 5, 2)$, $\mathbf{v} = (2, 4, -9)$

c. $\mathbf{u} = (1, 0, 1, 0)$, $\mathbf{v} = (-3, -3, -3, -3)$

2. **a.** $\mathbf{u} = (-1, 0)$, $\mathbf{v} = (3, 8)$

b. $\mathbf{u} = (4, 1, 8)$, $\mathbf{v} = (1, 0, -3)$

c. $\mathbf{u} = (2, 1, 7, -1)$, $\mathbf{v} = (4, 0, 0, 0)$

In Exercises 3–4, find the cosine of the angle between the vectors with respect to the standard inner product on P_2.

3. $\mathbf{p} = -1 + 5x + 2x^2$, $\mathbf{q} = 2 + 4x - 9x^2$

4. $\mathbf{p} = x - x^2$, $\mathbf{q} = 7 + 3x + 3x^2$

In Exercises 5–6, find the cosine of the angle between A and B with respect to the standard inner product on M_{22}.

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

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

In Exercises 7–8, determine whether the vectors are orthogonal with respect to the Euclidean inner product.

7. **a.** $\mathbf{u} = (-1, 3, 2)$, $\mathbf{v} = (4, 2, -1)$

b. $\mathbf{u} = (-2, -2, -2)$, $\mathbf{v} = (1, 1, 1)$

c. $\mathbf{u} = (a, b)$, $\mathbf{v} = (-b, a)$

8. **a.** $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (0, 0, 0)$

b. $\mathbf{u} = (-4, 6, -10, 1)$, $\mathbf{v} = (2, 1, -2, 9)$

c. $\mathbf{u} = (a, b, c)$, $\mathbf{v} = (-c, 0, a)$

In Exercises 9–10, show that the vectors are orthogonal with respect to the standard inner product on P_2.

9. $\mathbf{p} = -1 - x + 2x^2$, $\mathbf{q} = 2x + x^2$

10. $\mathbf{p} = 2 - 3x + x^2$, $\mathbf{q} = 4 + 2x - 2x^2$

In Exercises 11–12, show that the matrices are orthogonal with respect to the standard inner product on M_{22}.

11. $U = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}$, $V = \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix}$

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

In Exercises 13–14, show that the vectors are not orthogonal with respect to the Euclidean inner product on R^2, and then find a value of k for which the vectors are orthogonal with respect to the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + ku_2v_2$.

13. $\mathbf{u} = (1, 3)$, $\mathbf{v} = (2, -1)$

14. $\mathbf{u} = (2, -4)$, $\mathbf{v} = (0, 3)$

15. If the vectors $\mathbf{u} = (1, 2)$ and $\mathbf{v} = (2, -4)$ are orthogonal with respect to the weighted Euclidean inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = w_1u_1v_1 + w_2u_2v_2$$

what must be true of the weights w_1 and w_2?

16. Let R^4 have the Euclidean inner product. Find two unit vectors that are orthogonal to all three of the vectors $\mathbf{u} = (2, 1, -4, 0)$, $\mathbf{v} = (-1, -1, 2, 2)$, and $\mathbf{w} = (3, 2, 5, 4)$.

17. Do there exist scalars k and l such that the vectors

$$\mathbf{p}_1 = 2 + kx + 6x^2, \quad \mathbf{p}_2 = l + 5x + 3x^2, \quad \mathbf{p}_3 = 1 + 2x + 3x^2$$

are mutually orthogonal with respect to the standard inner product on P_2?

18. Show that the vectors

$$\mathbf{u} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 5 \\ -8 \end{bmatrix}$$

are orthogonal with respect to the inner product on R^2 that is generated by the matrix

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

[See Formulas (5) and (6) of Section 6.1.]

19. Let P_2 have the evaluation inner product at the points

$$x_0 = -2, \quad x_1 = 0, \quad x_2 = 2$$

Show that the vectors $\mathbf{p} = x$ and $\mathbf{q} = x^2$ are orthogonal with respect to this inner product.

20. Let M_{22} have the standard inner product. Determine whether the matrix A is in the subspace spanned by the matrices U and V.

$$A = \begin{bmatrix} -1 & 1 \\ 0 & 2 \end{bmatrix}, \quad U = \begin{bmatrix} 1 & -1 \\ 3 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} 4 & 0 \\ 9 & 2 \end{bmatrix}$$

In Exercises 21–24, confirm that the Cauchy–Schwarz inequality holds for the given vectors using the stated inner product.

21. $\mathbf{u} = (1, 0, 3)$, $\mathbf{v} = (2, 1, -1)$ using the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + 3u_2v_2 + u_3v_3$ in R^3.

22. $U = \begin{bmatrix} -1 & 2 \\ 6 & 1 \end{bmatrix}$ and $V = \begin{bmatrix} 1 & 0 \\ 3 & 3 \end{bmatrix}$

using the standard inner product on M_{22}.

23. $\mathbf{p} = -1 + 2x + x^2$ and $\mathbf{q} = 2 - 4x^2$ using the standard inner product on P_2.

24. The vectors

$$\mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

with respect to the inner product in Exercise 18.

25. Let R^4 have the Euclidean inner product, and suppose that $\mathbf{u} = (-1, 1, 0, 2)$. Determine whether the vector $\mathbf{u}$ is orthogonal to the subspace spanned by the vectors $\mathbf{w}_1 = (1, -1, 3, 0)$ and $\mathbf{w}_2 = (4, 0, 9, 2)$.

26. Let P_3 have the standard inner product, and let

$$\mathbf{p} = -1 - x + 2x^2 + 4x^3$$

Determine whether the polynomial $\mathbf{p}$ is orthogonal to the subspace spanned by the polynomials $\mathbf{w}_1 = 2 - x^2 + x^3$ and $\mathbf{w}_2 = 4x - 2x^2 + 2x^3$.

In Exercises 27–28, find a basis for the orthogonal complement of the subspace of R^n spanned by the vectors.

27. $\mathbf{v}_1 = (1, 4, 5, 2)$, $\mathbf{v}_2 = (2, 1, 3, 0)$, $\mathbf{v}_3 = (-1, 3, 2, 2)$

28. $\mathbf{v}_1 = (1, 4, 5, 6, 9)$, $\mathbf{v}_2 = (3, -2, 1, 4, -1)$, $\mathbf{v}_3 = (-1, 0, -1, -2, -1)$, $\mathbf{v}_4 = (2, 3, 5, 7, 8)$

In Exercises 29–30, assume that R^n has the Euclidean inner product.

29. **a.** Let W be the line in R^2 with equation $y = 2x$. Find an equation for $W^\perp$.

b. Let W be the plane in R^3 with equation $x - 2y - 3z = 0$. Find parametric equations for $W^\perp$.

30. **a.** Let W be the y-axis in an xyz-coordinate system in R^3. Describe the subspace $W^\perp$.

b. Let W be the yz-plane of an xyz-coordinate system in R^3. Describe the subspace $W^\perp$.

31. ***(Calculus required)*** Let $C[0, 1]$ have the integral inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_0^1 p(x)q(x)\,dx$$

and let $\mathbf{p} = p(x) = x$ and $\mathbf{q} = q(x) = x^2$.

a. Find $\langle \mathbf{p}, \mathbf{q} \rangle$.

b. Find $\|\mathbf{p}\|$ and $\|\mathbf{q}\|$.

32. **a.** Find the cosine of the angle between the vectors $\mathbf{p}$ and $\mathbf{q}$ in Exercise 31.

b. Find the distance between the vectors $\mathbf{p}$ and $\mathbf{q}$ in Exercise 31.

33. ***(Calculus required)*** Let $C[-1, 1]$ have the integral inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_{-1}^1 p(x)q(x)\,dx$$

and let $\mathbf{p} = p(x) = x^2 - x$ and $\mathbf{q} = q(x) = x + 1$.

a. Find $\langle \mathbf{p}, \mathbf{q} \rangle$.

b. Find $\|\mathbf{p}\|$ and $\|\mathbf{q}\|$.

34. **a.** Find the cosine of the angle between the vectors $\mathbf{p}$ and $\mathbf{q}$ in Exercise 33.

b. Find the distance between the vectors $\mathbf{p}$ and $\mathbf{q}$ in Exercise 33.

35. ***(Calculus required)*** Let $C[0, 1]$ have the inner product in Exercise 31.

a. Show that the vectors

$$\mathbf{p} = p(x) = 1 \quad \text{and} \quad \mathbf{q} = q(x) = \tfrac{1}{2} - x$$

are orthogonal.

b. Show that the vectors in part (a) satisfy the Theorem of Pythagoras.

36. ***(Calculus required)*** Let $C[-1, 1]$ have the inner product in Exercise 33.

a. Show that the vectors

$$\mathbf{p} = p(x) = x \quad \text{and} \quad \mathbf{q} = q(x) = x^2 - 1$$

are orthogonal.

b. Show that the vectors in part (a) satisfy the Theorem of Pythagoras.

37. Let V be an inner product space. Show that if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal unit vectors in V, then $\|\mathbf{u} - \mathbf{v}\| = \sqrt{2}$.

38. Let V be an inner product space. Show that if $\mathbf{w}$ is orthogonal to both $\mathbf{u}_1$ and $\mathbf{u}_2$, then it is orthogonal to $k_1\mathbf{u}_1 + k_2\mathbf{u}_2$ for all scalars k_1 and k_2. Interpret this result geometrically in the case where V is R^3 with the Euclidean inner product.

39. ***(Calculus required)*** Let $C[0, \pi]$ have the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_0^\pi f(x)g(x)\,dx$$

and let $\mathbf{f}_n = \cos nx$ $(n = 0, 1, 2, \ldots)$. Show that if $k \neq l$, then $\mathbf{f}_k$ and $\mathbf{f}_l$ are orthogonal vectors.

40. As shown in the figure below, the vectors $\mathbf{u} = (1, \sqrt{3})$ and $\mathbf{v} = (-1, \sqrt{3})$ have norm 2 and an angle of 60° between them relative to the Euclidean inner product. Find a weighted Euclidean inner product with respect to which $\mathbf{u}$ and $\mathbf{v}$ are orthogonal unit vectors.

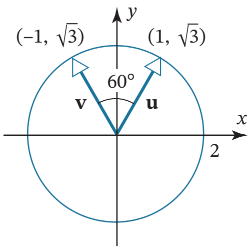


FIGURE Ex-40

Working with Proofs

41. Let V be an inner product space. Prove that if $\mathbf{w}$ is orthogonal to each of the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$, then it is orthogonal to every vector in span$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$.

42. Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a basis for an inner product space V. Prove that the zero vector is the only vector in V that is orthogonal to all of the basis vectors.

43. Let $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ be a basis for a subspace W of V. Prove that $W^\perp$ consists of all vectors in V that are orthogonal to every basis vector.

44. Prove the following generalization of Theorem 6.2.3: If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ are pairwise orthogonal vectors in an inner product space V, then

$$\|\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_r\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 + \cdots + \|\mathbf{v}_r\|^2$$

45. Prove: If $\mathbf{u}$ and $\mathbf{v}$ are $n \times 1$ matrices and A is an $n \times n$ matrix, then

$$(\mathbf{v}^T A^T A \mathbf{u})^2 \leq (\mathbf{u}^T A^T A \mathbf{u})(\mathbf{v}^T A^T A \mathbf{v})$$

46. Use the Cauchy–Schwarz inequality to prove that for all real values of a, b, and θ,

$$(a \cos\theta + b \sin\theta)^2 \leq a^2 + b^2$$

47. Prove: If $w_1, w_2, \ldots, w_n$ are positive real numbers, and if $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ are any two vectors in R^n, then

$$|w_1 u_1 v_1 + w_2 u_2 v_2 + \cdots + w_n u_n v_n| \leq$$
$$(w_1 u_1^2 + w_2 u_2^2 + \cdots + w_n u_n^2)^{1/2} (w_1 v_1^2 + w_2 v_2^2 + \cdots + w_n v_n^2)^{1/2}$$

48. Prove that equality holds in the Cauchy–Schwarz inequality if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent.

49. ***(Calculus required)*** Let $f(x)$ and $g(x)$ be continuous functions on $[0, 1]$. Prove:

a. $$\left[\int_0^1 f(x)g(x)\,dx\right]^2 \leq \left[\int_0^1 f^2(x)\,dx\right]\left[\int_0^1 g^2(x)\,dx\right]$$

b. $$\left[\int_0^1 [f(x)+g(x)]^2\,dx\right]^{1/2} \leq \left[\int_0^1 f^2(x)\,dx\right]^{1/2} + \left[\int_0^1 g^2(x)\,dx\right]^{1/2}$$

[*Hint:* Use the Cauchy–Schwarz inequality.]

50. Prove that Formula (4) holds for all nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in a real inner product space V.

51. Let $T_A : R^2 \to R^2$ be multiplication by

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

and let $\mathbf{x} = (1, 1)$.

a. Assuming that R^2 has the Euclidean inner product, find all vectors $\mathbf{v}$ in R^2 such that $\langle \mathbf{x}, \mathbf{v} \rangle = \langle T_A(\mathbf{x}), T_A(\mathbf{v}) \rangle$.

b. Assuming that R^2 has the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + 3u_2v_2$, find all vectors $\mathbf{v}$ in R^2 such that $\langle \mathbf{x}, \mathbf{v} \rangle = \langle T_A(\mathbf{x}), T_A(\mathbf{v}) \rangle$.

52. Let $T : P_2 \to P_2$ be the linear transformation defined by

$$T(a + bx + cx^2) = 3a - cx^2$$

and let $\mathbf{p} = 1 + x$.

a. Assuming that P_2 has the standard inner product, find all vectors $\mathbf{q}$ in P_2 such that $\langle \mathbf{p}, \mathbf{q} \rangle = \langle T(\mathbf{p}), T(\mathbf{q}) \rangle$.

b. Assuming that P_2 has the evaluation inner product at the points $x_0 = -1, x_1 = 0, x_2 = 1$, find all vectors $\mathbf{q}$ in P_2 such that $\langle \mathbf{p}, \mathbf{q} \rangle = \langle T(\mathbf{p}), T(\mathbf{q}) \rangle$.

True-False Exercises

TF. In parts **(a)–(f)** determine whether the statement is true or false, and justify your answer.

a. If $\mathbf{u}$ is orthogonal to every vector of a subspace W, then $\mathbf{u} = \mathbf{0}$.

b. If $\mathbf{u}$ is a vector in both W and $W^\perp$, then $\mathbf{u} = \mathbf{0}$.

c. If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $W^\perp$, then $\mathbf{u} + \mathbf{v}$ is in $W^\perp$.

d. If $\mathbf{u}$ is a vector in $W^\perp$ and k is a real number, then $k\mathbf{u}$ is in $W^\perp$.

e. If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, then $|\langle \mathbf{u}, \mathbf{v} \rangle| = \|\mathbf{u}\|\|\mathbf{v}\|$.

f. If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, then $\|\mathbf{u} + \mathbf{v}\| = \|\mathbf{u}\| + \|\mathbf{v}\|$.

Working with Technology

T1. a. We know that the row space and null space of a matrix are orthogonal complements relative to the Euclidean inner product. Confirm this fact for the matrix

$$A = \begin{bmatrix} 2 & -1 & 3 & 5 \\ 4 & -3 & 1 & 3 \\ 3 & -2 & 3 & 4 \\ 4 & -1 & 15 & 17 \\ 7 & -6 & -7 & 0 \end{bmatrix}$$

b. Find a basis for the orthogonal complement of the column space of A.

T2. In each part, confirm that the vectors $\mathbf{u}$ and $\mathbf{v}$ satisfy the Cauchy–Schwarz inequality relative to the stated inner product.

a. M_{44} with the standard inner product.

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

b. R^4 with the weighted Euclidean inner product with weights $w_1 = \frac{1}{2}, w_2 = \frac{1}{4}, w_3 = \frac{1}{8}, w_4 = \frac{1}{8}$.

$$\mathbf{u} = (1, -2, 2, 1) \quad \text{and} \quad \mathbf{v} = (0, -3, 3, -2)$$

6.3 Gram–Schmidt Process; *QR*-Decomposition

In many problems involving vector spaces, the problem solver is free to choose any basis for the vector space that seems appropriate. In inner product spaces, the solution of a problem can often be simplified by choosing a basis in which the vectors are orthogonal to one another. In this section we will show how such bases can be obtained.

Orthogonal and Orthonormal Sets

Recall from Section 6.2 that two vectors in an inner product space are said to be *orthogonal* if their inner product is zero. The following definition extends the notion of orthogonality to *sets* of vectors in an inner product space.

Definition 1

A set of two or more vectors in a real inner product space is said to be ***orthogonal*** if all pairs of distinct vectors in the set are orthogonal. An orthogonal set in which each vector has norm 1 is said to be ***orthonormal***.

EXAMPLE 1 | An Orthogonal Set in R^3

Let

$$\mathbf{v}_1 = (0, 1, 0), \quad \mathbf{v}_2 = (1, 0, 1), \quad \mathbf{v}_3 = (1, 0, -1)$$

and assume that R^3 has the Euclidean inner product. It follows that $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set since $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{v}_3 \rangle = \langle \mathbf{v}_2, \mathbf{v}_3 \rangle = 0$.

It frequently happens that one has found a set of orthogonal vectors in an inner product space but what is actually needed is a set of *orthonormal* vectors. A simple way to convert an orthogonal set of nonzero vectors into an orthonormal set is to multiply each vector $\mathbf{v}$ in the orthogonal set by the reciprocal of its length to create a vector of norm 1 (called a ***unit vector***). To see why this works, suppose that $\mathbf{v}$ is a nonzero vector in an inner product space, and let

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v} \tag{1}$$

Note that Formula (1) is identical to Formula (4) of Section 3.2, but whereas Formula (4) was valid only for vectors in R^n with the Euclidean inner product, Formula (1) is valid in general inner product spaces.

Then it follows from Theorem 6.1.1(b) with $k = \|\mathbf{v}\|$ that

$$\|\mathbf{u}\| = \left\| \frac{1}{\|\mathbf{v}\|}\mathbf{v} \right\| = \left| \frac{1}{\|\mathbf{v}\|} \right| \|\mathbf{v}\| = \frac{1}{\|\mathbf{v}\|}\|\mathbf{v}\| = 1$$

This process of multiplying $\mathbf{v}$ by the reciprocal of its length is called ***normalizing*** $\mathbf{v}$. We leave it as an exercise to show that normalizing the vectors in an orthogonal set of nonzero vectors preserves the orthogonality of the vectors and produces an orthonormal set.

EXAMPLE 2 | Constructing an Orthonormal Set

The Euclidean norms of the vectors in Example 1 are

$$\|\mathbf{v}_1\| = 1, \quad \|\mathbf{v}_2\| = \sqrt{2}, \quad \|\mathbf{v}_3\| = \sqrt{2}$$

Consequently, normalizing $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ yields

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = (0, 1, 0), \quad \mathbf{u}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right),$$

$$\mathbf{u}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right)$$

We leave it for you to verify that the set $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is orthonormal by showing that

$$\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = \langle \mathbf{u}_1, \mathbf{u}_3 \rangle = \langle \mathbf{u}_2, \mathbf{u}_3 \rangle = 0 \quad \text{and} \quad \|\mathbf{u}_1\| = \|\mathbf{u}_2\| = \|\mathbf{u}_3\| = 1$$

In R^2 any two nonzero perpendicular vectors are linearly independent because neither is a scalar multiple of the other; and in R^3 any three nonzero mutually perpendicular vectors are linearly independent because no one lies in the plane of the other two (and hence is not expressible as a linear combination of the other two). The following theorem generalizes these observations.

Theorem 6.3.1

If $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is an orthogonal set of nonzero vectors in an inner product space, then S is linearly independent.

Proof Assume that

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n = \mathbf{0} \tag{2}$$

To demonstrate that $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is linearly independent, we must prove that

$$k_1 = k_2 = \cdots = k_n = 0$$

For each $\mathbf{v}_i$ in S, it follows from (2) that

$$\langle k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n, \mathbf{v}_i \rangle = \langle \mathbf{0}, \mathbf{v}_i \rangle = 0$$

or, equivalently,

$$k_1\langle \mathbf{v}_1, \mathbf{v}_i \rangle + k_2\langle \mathbf{v}_2, \mathbf{v}_i \rangle + \cdots + k_n\langle \mathbf{v}_n, \mathbf{v}_i \rangle = 0$$

From the orthogonality of S it follows that $\langle \mathbf{v}_j, \mathbf{v}_i \rangle = 0$ when $j \neq i$, so this equation reduces to

$$k_i\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 0$$

Since the vectors in S are assumed to be nonzero, it follows from the positivity axiom for inner products that $\langle \mathbf{v}_i, \mathbf{v}_i \rangle \neq 0$. Thus, the preceding equation implies that each k_i in Equation (2) is zero, which is what we wanted to prove. ■

In an inner product space, a basis consisting of orthonormal vectors is called an ***orthonormal basis***, and a basis of orthogonal vectors is called an ***orthogonal basis***. A familiar example of an orthonormal basis is the standard basis for R^n with the Euclidean inner product:

$$\mathbf{e}_1 = (1, 0, 0, \dots, 0), \quad \mathbf{e}_2 = (0, 1, 0, \dots, 0), \dots, \quad \mathbf{e}_n = (0, 0, 0, \dots, 1)$$

Since an orthonormal set is orthogonal, and since its vectors are nonzero (norm 1), it follows from Theorem 6.3.1 that every *orthonormal* set is linearly independent.

EXAMPLE 3 | An Orthonormal Basis for P_n

Recall from Example 7 of Section 6.1 that the standard inner product of the polynomials

$$\mathbf{p} = a_0 + a_1x + \cdots + a_nx^n \quad \text{and} \quad \mathbf{q} = b_0 + b_1x + \cdots + b_nx^n$$

is

$$\langle \mathbf{p}, \mathbf{q} \rangle = a_0b_0 + a_1b_1 + \cdots + a_nb_n$$

and the norm of $\mathbf{p}$ relative to this inner product is

$$\|\mathbf{p}\| = \sqrt{\langle \mathbf{p}, \mathbf{p} \rangle} = \sqrt{a_0^2 + a_1^2 + \cdots + a_n^2}$$

Using these formulas you should be able to show that the standard basis

$$S = \{1, x, x^2, \dots, x^n\}$$

is orthonormal with respect to this inner product (verify).

EXAMPLE 4 | An Orthonormal Basis

In Example 2 we showed that the vectors

$$\mathbf{u}_1 = (0, 1, 0), \quad \mathbf{u}_2 = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right), \quad \text{and} \quad \mathbf{u}_3 = \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right)$$

form an orthonormal set with respect to the Euclidean inner product on R^3. By Theorem 6.3.1, these vectors form a linearly independent set, and since R^3 is three-dimensional, it follows from Theorem 4.6.4 that $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis for R^3.

Coordinates Relative to Orthonormal Bases

One way to express a vector $\mathbf{u}$ as a linear combination of basis vectors

$$S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$$

is to convert the vector equation

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

to a linear system and solve for the coefficients $c_1, c_2, \dots, c_n$. However, if the basis happens to be orthogonal or orthonormal, then the following theorem shows that the coefficients can be obtained more simply by computing appropriate inner products.

Theorem 6.3.2

(*a*) If $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is an orthogonal basis for an inner product space V, and if $\mathbf{u}$ is any vector in V, then

$$\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_n \rangle}{\|\mathbf{v}_n\|^2}\mathbf{v}_n \tag{3}$$

(*b*) If $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is an orthonormal basis for an inner product space V, and if $\mathbf{u}$ is any vector in V, then

$$\mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle \mathbf{v}_2 + \cdots + \langle \mathbf{u}, \mathbf{v}_n \rangle \mathbf{v}_n \tag{4}$$

Proof (a) Since $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is a basis for V, every vector $\mathbf{u}$ in V can be expressed in the form

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

We will complete the proof by showing that

$$c_i = \frac{\langle \mathbf{u}, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|^2} \tag{5}$$

for $i = 1, 2, \dots, n$. To do this, observe first that

$$\begin{aligned}\langle \mathbf{u}, \mathbf{v}_i \rangle &= \langle c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n, \mathbf{v}_i \rangle \\ &= c_1\langle \mathbf{v}_1, \mathbf{v}_i \rangle + c_2\langle \mathbf{v}_2, \mathbf{v}_i \rangle + \cdots + c_n\langle \mathbf{v}_n, \mathbf{v}_i \rangle\end{aligned}$$

Since S is an orthogonal set, all of the inner products in the last equality are zero except the ith, so we have

$$\langle \mathbf{u}, \mathbf{v}_i \rangle = c_i\langle \mathbf{v}_i, \mathbf{v}_i \rangle = c_i\|\mathbf{v}_i\|^2$$

Solving this equation for c_i yields (5), which completes the proof.

Proof (b) In this case, $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \cdots = \|\mathbf{v}_n\| = 1$, so Formula (3) simplifies to Formula (4). ■

Using the terminology and notation from Definition 2 of Section 4.5, it follows from Theorem 6.3.2 that the coordinate vector of a vector $\mathbf{u}$ in V relative to an orthogonal basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is

$$\boxed{(\mathbf{u})_S = \left(\frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}, \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}, \dots, \frac{\langle \mathbf{u}, \mathbf{v}_n \rangle}{\|\mathbf{v}_n\|^2}\right)} \tag{6}$$

and relative to an orthonormal basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is

$$\boxed{(\mathbf{u})_S = \big(\langle \mathbf{u}, \mathbf{v}_1 \rangle, \langle \mathbf{u}, \mathbf{v}_2 \rangle, \dots, \langle \mathbf{u}, \mathbf{v}_n \rangle\big)} \tag{7}$$

EXAMPLE 5 | A Coordinate Vector Relative to an Orthonormal Basis

Let

$$\mathbf{v}_1 = (0, 1, 0), \quad \mathbf{v}_2 = \left(-\tfrac{4}{5}, 0, \tfrac{3}{5}\right), \quad \mathbf{v}_3 = \left(\tfrac{3}{5}, 0, \tfrac{4}{5}\right)$$

It is easy to check that $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis for R^3 with the Euclidean inner product. Express the vector $\mathbf{u} = (1, 1, 1)$ as a linear combination of the vectors in S, and find the coordinate vector $(\mathbf{u})_S$.

Solution We leave it for you to verify that

$$\langle \mathbf{u}, \mathbf{v}_1 \rangle = 1, \quad \langle \mathbf{u}, \mathbf{v}_2 \rangle = -\tfrac{1}{5}, \quad \text{and} \quad \langle \mathbf{u}, \mathbf{v}_3 \rangle = \tfrac{7}{5}$$

Therefore, by Theorem 6.3.2 we have

$$\mathbf{u} = \mathbf{v}_1 - \tfrac{1}{5}\mathbf{v}_2 + \tfrac{7}{5}\mathbf{v}_3$$

that is,

$$(1, 1, 1) = (0, 1, 0) - \tfrac{1}{5}\left(-\tfrac{4}{5}, 0, \tfrac{3}{5}\right) + \tfrac{7}{5}\left(\tfrac{3}{5}, 0, \tfrac{4}{5}\right)$$

Thus, the coordinate vector of $\mathbf{u}$ relative to S is

$$(\mathbf{u})_S = \left(\langle \mathbf{u}, \mathbf{v}_1\rangle, \langle \mathbf{u}, \mathbf{v}_2\rangle, \langle \mathbf{u}, \mathbf{v}_3\rangle\right) = \left(1, -\tfrac{1}{5}, \tfrac{7}{5}\right)$$

EXAMPLE 6 | An Orthonormal Basis from an Orthogonal Basis

(*a*) Show that the vectors

$$\mathbf{w}_1 = (0, 2, 0), \quad \mathbf{w}_2 = (3, 0, 3), \quad \mathbf{w}_3 = (-4, 0, 4)$$

form an orthogonal basis for R^3 with the Euclidean inner product, and use that basis to find an orthonormal basis by normalizing each vector.

(*b*) Express the vector $\mathbf{u} = (1, 2, 4)$ as a linear combination of the orthonormal basis vectors obtained in part (*a*).

Solution (*a*) The given vectors form an orthogonal set since

$$\langle \mathbf{w}_1, \mathbf{w}_2\rangle = 0, \quad \langle \mathbf{w}_1, \mathbf{w}_3\rangle = 0, \quad \langle \mathbf{w}_2, \mathbf{w}_3\rangle = 0$$

It follows from Theorem 6.3.1 that these vectors are linearly independent and hence form a basis for R^3 by Theorem 4.6.4. We leave it for you to calculate the norms of $\mathbf{w}_1, \mathbf{w}_2$, and $\mathbf{w}_3$ and then obtain the orthonormal basis

$$\mathbf{v}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} = (0, 1, 0), \quad \mathbf{v}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right),$$

$$\mathbf{v}_3 = \frac{\mathbf{w}_3}{\|\mathbf{w}_3\|} = \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$$

Solution (*b*) It follows from Formula (4) that

$$\mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1\rangle \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2\rangle \mathbf{v}_2 + \langle \mathbf{u}, \mathbf{v}_3\rangle \mathbf{v}_3$$

We leave it for you to confirm that

$$\langle \mathbf{u}, \mathbf{v}_1\rangle = (1, 2, 4) \cdot (0, 1, 0) = 2$$

$$\langle \mathbf{u}, \mathbf{v}_2\rangle = (1, 2, 4) \cdot \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) = \frac{5}{\sqrt{2}}$$

$$\langle \mathbf{u}, \mathbf{v}_3\rangle = (1, 2, 4) \cdot \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) = \frac{3}{\sqrt{2}}$$

and hence that

$$(1, 2, 4) = 2(0, 1, 0) + \frac{5}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) + \frac{3}{\sqrt{2}}\left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$$

Orthogonal Projections

Many applied problems are best solved by working with orthogonal or orthonormal basis vectors. Such bases are typically found by starting with some simple basis (say a standard basis) and then converting that basis into an orthogonal or orthonormal basis. To explain exactly how that is done will require some preliminary ideas about orthogonal projections.

In Section 3.3 we proved a result called the *Projection Theorem* (see Theorem 3.3.2) that dealt with the problem of decomposing a vector $\mathbf{u}$ in R^n into a sum of two terms, $\mathbf{w}_1$ and $\mathbf{w}_2$, in which $\mathbf{w}_1$ is the orthogonal projection of $\mathbf{u}$ on some nonzero vector $\mathbf{a}$ and $\mathbf{w}_2$ is orthogonal to $\mathbf{w}_1$ (Figure 3.3.2). That result is a special case of the following more general theorem, which we will state without proof.

Theorem 6.3.3

Projection Theorem

If W is a finite-dimensional subspace of an inner product space V, then every vector $\mathbf{u}$ in V can be expressed in exactly one way as

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2 \tag{8}$$

where $\mathbf{w}_1$ is in W and $\mathbf{w}_2$ is in $W^\perp$.

The vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ in Formula (8) are commonly denoted by

$$\mathbf{w}_1 = \text{proj}_W \mathbf{u} \quad \text{and} \quad \mathbf{w}_2 = \text{proj}_{W^\perp} \mathbf{u} \tag{9}$$

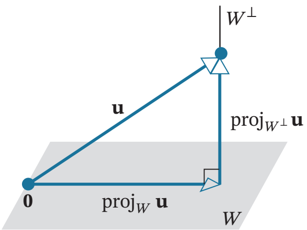


FIGURE 6.3.1

These are called the ***orthogonal projection of*** $\mathbf{u}$ ***on*** W and the ***orthogonal projection of*** $\mathbf{u}$ ***on*** $W^\perp$, respectively. The vector $\mathbf{w}_2$ is also called the ***component of*** $\mathbf{u}$ ***orthogonal to*** W. Using the notation in (9), Formula (8) can be expressed as

$$\boxed{\mathbf{u} = \text{proj}_W \mathbf{u} + \text{proj}_{W^\perp} \mathbf{u}} \tag{10}$$

(**Figure 6.3.1**). Moreover, since $\text{proj}_{W^\perp}\mathbf{u} = \mathbf{u} - \text{proj}_W \mathbf{u}$, we can also express Formula (10) as

$$\boxed{\mathbf{u} = \text{proj}_W \mathbf{u} + (\mathbf{u} - \text{proj}_W \mathbf{u})} \tag{11}$$

The following theorem provides formulas for calculating orthogonal projections.

Theorem 6.3.4

Let W be a finite-dimensional subspace of an inner product space V.

(*a*) If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is an orthogonal basis for W, and $\mathbf{u}$ is any vector in V, then

$$\text{proj}_W \mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_r\rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r \tag{12}$$

(*b*) If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is an orthonormal basis for W, and $\mathbf{u}$ is any vector in V, then

$$\text{proj}_W \mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1\rangle\mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2\rangle\mathbf{v}_2 + \cdots + \langle \mathbf{u}, \mathbf{v}_r\rangle\mathbf{v}_r \tag{13}$$

Although Formulas (12) and (13) are expressed in terms of orthogonal and orthonormal basis vectors, the resulting vector $\text{proj}_W \mathbf{u}$ does not depend on the basis vectors that are used.

Proof (a) It follows from Theorem 6.3.3 that the vector $\mathbf{u}$ can be expressed in the form $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1 = \text{proj}_W \mathbf{u}$ is in W and $\mathbf{w}_2$ is in $W^\perp$; and it follows from Theorem 6.3.2 that the component $\text{proj}_W \mathbf{u} = \mathbf{w}_1$ can be expressed in terms of the basis vectors for W as

$$\text{proj}_W \mathbf{u} = \mathbf{w}_1 = \frac{\langle \mathbf{w}_1, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{w}_1, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{w}_1, \mathbf{v}_r\rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r \tag{14}$$

Since $\mathbf{w}_2$ is orthogonal to W, it follows that

$$\langle \mathbf{w}_2, \mathbf{v}_1\rangle = \langle \mathbf{w}_2, \mathbf{v}_2\rangle = \cdots = \langle \mathbf{w}_2, \mathbf{v}_r\rangle = 0$$

so we can rewrite (14) as

$$\text{proj}_W \mathbf{u} = \mathbf{w}_1 = \frac{\langle \mathbf{w}_1 + \mathbf{w}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{w}_1 + \mathbf{w}_2, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{w}_1 + \mathbf{w}_2, \mathbf{v}_r \rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r$$

or, equivalently, as

$$\text{proj}_W \mathbf{u} = \mathbf{w}_1 = \frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_r \rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r$$

Proof (b) In this case, $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \cdots = \|\mathbf{v}_r\| = 1$, so Formula (14) simplifies to Formula (13). ■

EXAMPLE 7 | Calculating Projections

Let R^3 have the Euclidean inner product, and let W be the subspace spanned by the orthonormal vectors $\mathbf{v}_1 = (0, 1, 0)$ and $\mathbf{v}_2 = \left(-\frac{4}{5}, 0, \frac{3}{5}\right)$. From Formula (13) the orthogonal projection of $\mathbf{u} = (1, 1, 1)$ on W is

$$\begin{aligned} \text{proj}_W \mathbf{u} &= \langle \mathbf{u}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle \mathbf{v}_2 \\ &= (1)(0, 1, 0) + \left(-\tfrac{1}{5}\right)\left(-\tfrac{4}{5}, 0, \tfrac{3}{5}\right) \\ &= \left(\tfrac{4}{25}, 1, -\tfrac{3}{25}\right) \end{aligned}$$

The component of $\mathbf{u}$ orthogonal to W is

$$\text{proj}_{W^\perp} \mathbf{u} = \mathbf{u} - \text{proj}_W \mathbf{u} = (1, 1, 1) - \left(\tfrac{4}{25}, 1, -\tfrac{3}{25}\right) = \left(\tfrac{21}{25}, 0, \tfrac{28}{25}\right)$$

Observe that $\text{proj}_{W^\perp} \mathbf{u}$ is orthogonal to both $\mathbf{v}_1$ and $\mathbf{v}_2$, so this vector is orthogonal to each vector in the space W spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$, as it should be.

A Geometric Interpretation of Orthogonal Projections

It follows from Formula (10) of Section 3.3 that each term in Formula (12) can be viewed as the orthogonal projection of $\mathbf{u}$ onto a 1-dimensional subspace. The first term is the orthogonal projection onto span$\{\mathbf{v}_1\}$, the second is the orthogonal projection onto span$\{\mathbf{v}_2\}$, and so forth. This suggests that we can think of (12) as the sum of orthogonal projections onto "axes" determined by the basis vectors for the subspace W (**Figure 6.3.2**).

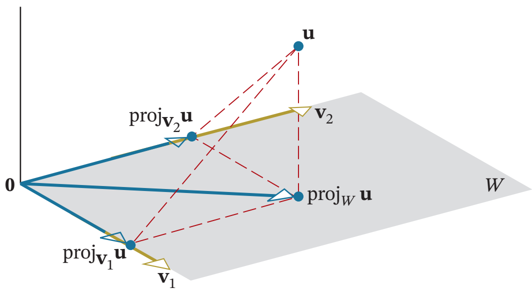


FIGURE 6.3.2

The Gram–Schmidt Process

We have seen that orthonormal bases exhibit a variety of useful properties. Our next theorem, which is the main result in this section, shows that every nonzero finite-dimensional vector space has an orthonormal basis. The proof of this result is extremely important since it provides an algorithm, or method, for converting an arbitrary basis into an orthonormal basis.

Theorem 6.3.5

Every nonzero finite-dimensional inner product space has an orthonormal basis.

Proof Let W be any nonzero finite-dimensional subspace of an inner product space, and suppose that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ is any basis for W. It suffices to show that W has an orthogonal basis since the vectors in that basis can be normalized to obtain an orthonormal basis. The following sequence of steps will produce an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ for W:

Step 1. Let $\mathbf{v}_1 = \mathbf{u}_1$.

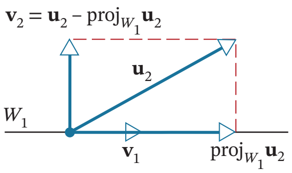


FIGURE 6.3.3

Step 2. As illustrated in **Figure 6.3.3**, we can obtain a vector $\mathbf{v}_2$ that is orthogonal to $\mathbf{v}_1$ by computing the component of $\mathbf{u}_2$ that is orthogonal to the space W_1 spanned by $\mathbf{v}_1$. Using Formula (12) to perform this computation, we obtain

$$\mathbf{v}_2 = \mathbf{u}_2 - \text{proj}_{W_1}\mathbf{u}_2 = \mathbf{u}_2 - \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1$$

Of course, if $\mathbf{v}_2 = \mathbf{0}$, then $\mathbf{v}_2$ is not a basis vector. But this cannot happen, since it would then follow from the preceding formula for $\mathbf{v}_2$ that

$$\mathbf{u}_2 = \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 = \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{u}_1\|^2}\mathbf{u}_1$$

which implies that $\mathbf{u}_2$ is a multiple of $\mathbf{u}_1$, contradicting the linear independence of the basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$.

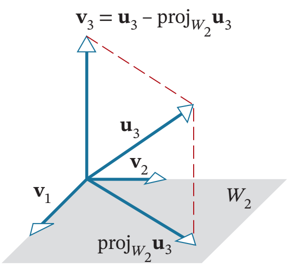


FIGURE 6.3.4

Step 3. To construct a vector $\mathbf{v}_3$ that is orthogonal to both $\mathbf{v}_1$ and $\mathbf{v}_2$, we compute the component of $\mathbf{u}_3$ orthogonal to the space W_2 spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$ (**Figure 6.3.4**). Using Formula (12) to perform this computation, we obtain

$$\mathbf{v}_3 = \mathbf{u}_3 - \text{proj}_{W_2}\mathbf{u}_3 = \mathbf{u}_3 - \frac{\langle \mathbf{u}_3, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_3, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2$$

As in Step 2, the linear independence of $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ ensures that $\mathbf{v}_3 \neq \mathbf{0}$. We leave the details for you.

Step 4. To determine a vector $\mathbf{v}_4$ that is orthogonal to $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$, we compute the component of $\mathbf{u}_4$ orthogonal to the space W_3 spanned by $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$. From (12),

$$\mathbf{v}_4 = \mathbf{u}_4 - \text{proj}_{W_3}\mathbf{u}_4 = \mathbf{u}_4 - \frac{\langle \mathbf{u}_4, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_4, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 - \frac{\langle \mathbf{u}_4, \mathbf{v}_3 \rangle}{\|\mathbf{v}_3\|^2}\mathbf{v}_3$$

Continuing in this way we will produce after r steps an orthogonal set of nonzero vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$. Since such sets are linearly independent, we will have produced an orthogonal basis for the r-dimensional space W. By normalizing these basis vectors we can obtain an orthonormal basis. ■

The step-by-step construction of an orthogonal (or orthonormal) basis given in the foregoing proof is called the ***Gram–Schmidt process***. For reference, we provide the following summary of the steps.

The Gram–Schmidt Process

To convert a basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ into an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$, perform the following computations:

Step 1. $\mathbf{v}_1 = \mathbf{u}_1$

Step 2. $\mathbf{v}_2 = \mathbf{u}_2 - \dfrac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1$

Step 3. $\mathbf{v}_3 = \mathbf{u}_3 - \dfrac{\langle \mathbf{u}_3, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \dfrac{\langle \mathbf{u}_3, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2$

Step 4. $\mathbf{v}_4 = \mathbf{u}_4 - \dfrac{\langle \mathbf{u}_4, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \dfrac{\langle \mathbf{u}_4, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 - \dfrac{\langle \mathbf{u}_4, \mathbf{v}_3 \rangle}{\|\mathbf{v}_3\|^2}\mathbf{v}_3$

$\vdots$

(continue for r steps)

Optional Step. To convert the orthogonal basis into an orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_r\}$, normalize the orthogonal basis vectors.

Historical Note

Jorgen Pederson Gram (1850–1916)

Gram was a Danish actuary whose early education was at village schools supplemented by private tutoring. He obtained a doctorate degree in mathematics while working for the Hafnia Life Insurance Company, where he specialized in the mathematics of accident insurance. It was in his dissertation that his contributions to the Gram–Schmidt process were formulated. He eventually became interested in abstract mathematics and received a gold medal from the Royal Danish Society of Sciences and Letters in recognition of his work. His lifelong interest in applied mathematics never wavered, however, and he produced a variety of treatises on Danish forest management.

Oswald Johannes Erhardt Schmidt (1875–1959)

Erhardt Schmidt was a German mathematician who studied for his doctoral degree at Göttingen University under David Hilbert, one of the giants of modern mathematics. For most of his life he taught at Berlin University where, in addition to making important contributions to many branches of mathematics, he fashioned some of Hilbert's ideas into a general concept, called a *Hilbert space*—a fundamental structure in the study of infinite-dimensional vector spaces. He first described the process that bears his name in a paper on integral equations that he published in 1907.

[Images: https://commons.wikimedia.org/wiki/Category:J%C3%B8rgen_Pedersen_Gram#/media/File:Jorgen_Gram.jpg. Public Domain. (Gram); Archives of the Mathematisches Forschungsinstitut Oberwolfach (Erhardt Schmidt)]

EXAMPLE 8 | Using the Gram–Schmidt Process

Assume that the vector space R^3 has the Euclidean inner product. Apply the Gram–Schmidt process to transform the basis vectors

$$\mathbf{u}_1 = (1, 1, 1), \quad \mathbf{u}_2 = (0, 1, 1), \quad \mathbf{u}_3 = (0, 0, 1)$$

into an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, and then normalize the orthogonal basis vectors to obtain an orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$.

Solution

Step 1. $\mathbf{v}_1 = \mathbf{u}_1 = (1, 1, 1)$

Step 2. $\mathbf{v}_2 = \mathbf{u}_2 - \text{proj}_{W_1}\mathbf{u}_2 = \mathbf{u}_2 - \dfrac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1$

$$= (0, 1, 1) - \frac{2}{3}(1, 1, 1) = \left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

Step 3. $\mathbf{v}_3 = \mathbf{u}_3 - \text{proj}_{W_2}\mathbf{u}_3 = \mathbf{u}_3 - \dfrac{\langle \mathbf{u}_3, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \dfrac{\langle \mathbf{u}_3, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2$

$$= (0, 0, 1) - \frac{1}{3}(1, 1, 1) - \frac{1/3}{2/3}\left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

$$= \left(0, -\frac{1}{2}, \frac{1}{2}\right)$$

Thus,

$$\mathbf{v}_1 = (1, 1, 1), \quad \mathbf{v}_2 = \left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right), \quad \mathbf{v}_3 = \left(0, -\frac{1}{2}, \frac{1}{2}\right)$$

form an orthogonal basis for R^3. The norms of these vectors are

$$\|\mathbf{v}_1\| = \sqrt{3}, \quad \|\mathbf{v}_2\| = \frac{\sqrt{6}}{3}, \quad \|\mathbf{v}_3\| = \frac{1}{\sqrt{2}}$$

so an orthonormal basis for R^3 is

$$\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \quad \mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \left(-\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right),$$

$$\mathbf{q}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

Remark In the last example we normalized at the end to convert the orthogonal basis into an orthonormal basis. Alternatively, we could have normalized each orthogonal basis vector as soon as it was obtained, thereby producing an orthonormal basis step by step. However, that procedure generally has the disadvantage in hand calculation of producing more square roots to manipulate. A more useful variation is to "scale" the orthogonal basis vectors at each step to eliminate some of the fractions. For example, after Step 2 above, we could have multiplied by 3 to produce $(-2, 1, 1)$ as the second orthogonal basis vector, thereby simplifying the calculations in Step 3.

CALCULUS REQUIRED

EXAMPLE 9 | Legendre Polynomials

Let the vector space P_2 have the inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_{-1}^{1} p(x)q(x)\,dx$$

Apply the Gram–Schmidt process to transform the standard basis $\{1, x, x^2\}$ for P_2 into an orthogonal basis $\{\phi_1(x), \phi_2(x), \phi_3(x)\}$.

Solution Take $\mathbf{u}_1 = 1$, $\mathbf{u}_2 = x$, and $\mathbf{u}_3 = x^2$.

Step 1. $\mathbf{v}_1 = \mathbf{u}_1 = 1$

Step 2. We have

$$\langle \mathbf{u}_2, \mathbf{v}_1 \rangle = \int_{-1}^{1} x\,dx = 0$$

so

$$\mathbf{v}_2 = \mathbf{u}_2 - \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 = \mathbf{u}_2 = x$$

Step 3. We have

$$\langle \mathbf{u}_3, \mathbf{v}_1 \rangle = \int_{-1}^{1} x^2\,dx = \frac{x^3}{3}\bigg]_{-1}^{1} = \frac{2}{3}$$

$$\langle \mathbf{u}_3, \mathbf{v}_2 \rangle = \int_{-1}^{1} x^3\,dx = \frac{x^4}{4}\bigg]_{-1}^{1} = 0$$

$$\|\mathbf{v}_1\|^2 = \langle \mathbf{v}_1, \mathbf{v}_1 \rangle = \int_{-1}^{1} 1\,dx = x\bigg]_{-1}^{1} = 2$$

so

$$\mathbf{v}_3 = \mathbf{u}_3 - \frac{\langle \mathbf{u}_3, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_3, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 = x^2 - \frac{1}{3}$$

Thus, we have obtained the orthogonal basis $\{\phi_1(x), \phi_2(x), \phi_3(x)\}$ in which

$$\phi_1(x) = 1, \quad \phi_2(x) = x, \quad \phi_3(x) = x^2 - \frac{1}{3}$$

Remark The orthogonal basis vectors in the last example are often scaled so all three functions have a value of 1 at $x = 1$. The resulting polynomials

$$1, \quad x, \quad \frac{1}{2}(3x^2 - 1)$$

which are known as the first three ***Legendre polynomials***, play an important role in a variety of applications. The scaling does not affect the orthogonality.

Extending Orthonormal Sets to Orthonormal Bases

Recall from part (b) of Theorem 4.6.5 that a linearly independent set in a finite-dimensional vector space can be enlarged to a basis by adding appropriate vectors. The following theorem is an analog of that result for orthogonal and orthonormal sets in finite-dimensional inner product spaces.

Theorem 6.3.6

If W is a finite-dimensional inner product space, then:

(a) Every orthogonal set of nonzero vectors in W can be enlarged to an orthogonal basis for W.

(b) Every orthonormal set in W can be enlarged to an orthonormal basis for W.

We will prove part (b) and leave part (a) as an exercise.

Proof (b) Suppose that $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s\}$ is an orthonormal set of vectors in W. Part (b) of Theorem 4.6.5 tells us that we can enlarge S to some basis

$$S' = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s, \mathbf{v}_{s+1}, \dots, \mathbf{v}_k\}$$

for W. If we now apply the Gram–Schmidt process to the set S', then the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s$ will not be affected since they are already orthonormal, and the resulting set

$$S'' = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s, \mathbf{v}'_{s+1}, \dots, \mathbf{v}'_k\}$$

will be an orthonormal basis for W. ■

OPTIONAL: *QR*-Decomposition

In recent years a numerical algorithm based on the Gram–Schmidt process, and known as ***QR-decomposition***, has assumed growing importance as the mathematical foundation for a wide variety of numerical algorithms, including those for computing eigenvalues of large matrices. The technical aspects of such algorithms are discussed in books that specialize in the numerical aspects of linear algebra. However, we will discuss some of the underlying ideas here. We begin by posing the following problem.

> **Problem** If A is an $m \times n$ matrix with linearly independent column vectors, and if Q is the matrix that results by applying the Gram–Schmidt process to the column vectors of A, what relationship, if any, exists between A and Q?

To solve this problem, suppose that the column vectors of A are $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ and that Q has orthonormal column vectors $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$. Thus, A and Q can be written in partitioned form as

$$A = [\mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_n] \quad \text{and} \quad Q = [\mathbf{q}_1 \mid \mathbf{q}_2 \mid \cdots \mid \mathbf{q}_n]$$

It follows from Theorem 6.3.2(b) that $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ are expressible in terms of the vectors $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$ as

$$\begin{aligned}
\mathbf{u}_1 &= \langle \mathbf{u}_1, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_1, \mathbf{q}_2 \rangle \mathbf{q}_2 + \cdots + \langle \mathbf{u}_1, \mathbf{q}_n \rangle \mathbf{q}_n \\
\mathbf{u}_2 &= \langle \mathbf{u}_2, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_2, \mathbf{q}_2 \rangle \mathbf{q}_2 + \cdots + \langle \mathbf{u}_2, \mathbf{q}_n \rangle \mathbf{q}_n \\
&\ \ \vdots \\
\mathbf{u}_n &= \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \mathbf{q}_2 + \cdots + \langle \mathbf{u}_n, \mathbf{q}_n \rangle \mathbf{q}_n
\end{aligned}$$

Recalling from Section 1.3 (Example 9) that the jth column vector of a matrix product is a linear combination of the column vectors of the first factor with coefficients coming from the jth column of the second factor, it follows that these relationships can be expressed in matrix form as

$$[\mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_n] = [\mathbf{q}_1 \mid \mathbf{q}_2 \mid \cdots \mid \mathbf{q}_n] \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ \langle \mathbf{u}_1, \mathbf{q}_2 \rangle & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & & \vdots \\ \langle \mathbf{u}_1, \mathbf{q}_n \rangle & \langle \mathbf{u}_2, \mathbf{q}_n \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{bmatrix}$$

or more briefly as

$$A = QR \tag{15}$$

where R is the second factor in the product. However, it is a property of the Gram–Schmidt process that for $j \ge 2$, the vector $\mathbf{q}_j$ is orthogonal to $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{j-1}$. Thus, all entries below the main diagonal of R are zero, and R has the form

$$R = \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ 0 & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{bmatrix} \tag{16}$$

We leave it for you to show that R is invertible by showing that its diagonal entries are nonzero. Thus, Equation (15) is a factorization of A into the product of a matrix Q with orthonormal column vectors and an invertible upper triangular matrix R. We call Equation (15) a ***QR-decomposition of A***. In summary, we have the following theorem.

Theorem 6.3.7

***QR*-Decomposition**

If A is an $m \times n$ matrix with linearly independent column vectors, then A can be factored as

$$A = QR$$

where Q is an $m \times n$ matrix with orthonormal column vectors, and R is an $n \times n$ invertible upper triangular matrix.

It is common in numerical linear algebra to say that a matrix with linearly independent columns has ***full column rank***.

Recall from Theorem 5.1.5 (the Equivalence Theorem) that a *square* matrix has linearly independent column vectors if and only if it is invertible. Thus, it follows from Theorem 6.3.7 that *every invertible matrix has a QR-decomposition*.

EXAMPLE 10 | QR-Decomposition of a 3×3 Matrix

Find a QR-decomposition of

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

Solution The column vectors of A are

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Applying the Gram–Schmidt process with normalization to these column vectors yields the orthonormal vectors (see Example 8)

$$\mathbf{q}_1 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \quad \mathbf{q}_2 = \begin{bmatrix} -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \quad \mathbf{q}_3 = \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Thus, it follows from Formula (16) that R is

$$R = \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \langle \mathbf{u}_3, \mathbf{q}_1 \rangle \\ 0 & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \langle \mathbf{u}_3, \mathbf{q}_2 \rangle \\ 0 & 0 & \langle \mathbf{u}_3, \mathbf{q}_3 \rangle \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{3}} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

from which it follows that a QR-decomposition of A is

$$\underset{A}{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}} = \underset{Q}{\begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{bmatrix}} \underset{R}{\begin{bmatrix} \frac{3}{\sqrt{3}} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}}$$

Exercise Set 6.3

1. In each part, determine whether the set of vectors is orthogonal and whether it is orthonormal with respect to the Euclidean inner product on R^2.

a. $(0,1),\ (2,0)$

b. $\left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right),\ \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

c. $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right),\ \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

d. $(0,0),\ (0,1)$

2. In each part, determine whether the set of vectors is orthogonal and whether it is orthonormal with respect to the Euclidean inner product on R^3.

a. $\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right),\ \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right),\ \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$

b. $\left(\frac{2}{3}, -\frac{2}{3}, \frac{1}{3}\right),\ \left(\frac{2}{3}, \frac{1}{3}, -\frac{2}{3}\right),\ \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right)$

c. $(1,0,0),\ \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right),\ (0,0,1)$

d. $\left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right),\ \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right)$

3. In each part, determine whether the set of vectors is orthogonal with respect to the standard inner product on P_2 (see Example 7 of Section 6.1).

a. $p_1(x) = \frac{2}{3} - \frac{2}{3}x + \frac{1}{3}x^2,\ p_2(x) = \frac{2}{3} + \frac{1}{3}x - \frac{2}{3}x^2,$
$p_3(x) = \frac{1}{3} + \frac{2}{3}x + \frac{2}{3}x^2$

b. $p_1(x) = 1,\ p_2(x) = \frac{1}{\sqrt{2}}x + \frac{1}{\sqrt{2}}x^2,\ p_3(x) = x^2$

4. In each part, determine whether the set of vectors is orthogonal with respect to the standard inner product on M_{22} (see Example 6 of Section 6.1).

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

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

In Exercises 5–6, show that the column vectors of A form an orthogonal basis for the column space of A with respect to the Euclidean inner product, and then find an orthonormal basis for that column space.

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

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

7. Verify that the vectors

$$\mathbf{v}_1 = \left(-\frac{3}{5}, \frac{4}{5}, 0\right),\ \mathbf{v}_2 = \left(\frac{4}{5}, \frac{3}{5}, 0\right),\ \mathbf{v}_3 = (0,0,1)$$

form an orthonormal basis for R^3 with respect to the Euclidean inner product, and then use Theorem 6.3.2(b) to express the vector $\mathbf{u} = (1,-2,2)$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$.

8. Use Theorem 6.3.2(b) to express the vector $\mathbf{u} = (3,-7,4)$ as a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$ in Exercise 7.

9. Verify that the vectors

$$\mathbf{v}_1 = (2,-2,1),\quad \mathbf{v}_2 = (2,1,-2),\quad \mathbf{v}_3 = (1,2,2)$$

form an orthogonal basis for R^3 with respect to the Euclidean inner product, and then use Theorem 6.3.2(a) to express the vector $\mathbf{u} = (-1,0,2)$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$.

10. Verify that the vectors

$$\mathbf{v}_1 = (1,-1,2,-1),\quad \mathbf{v}_2 = (-2,2,3,2),$$
$$\mathbf{v}_3 = (1,2,0,-1),\quad \mathbf{v}_4 = (1,0,0,1)$$

form an orthogonal basis for R^4 with respect to the Euclidean inner product, and then use Theorem 6.3.2(a) to express the vector $\mathbf{u} = (1,1,1,1)$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, and $\mathbf{v}_4$.

In Exercises 11–14, find the coordinate vector $(\mathbf{u})_S$ for the vector $\mathbf{u}$ and the basis S that were given in the stated exercise.

11. Exercise 7

12. Exercise 8

13. Exercise 9

14. Exercise 10

In Exercises 15–18, let R^2 have the Euclidean inner product.

a. *Find the orthogonal projection of $\mathbf{u}$ onto the line spanned by the vector $\mathbf{v}$.*

b. *Find the component of $\mathbf{u}$ orthogonal to the line spanned by the vector $\mathbf{v}$, and confirm that this component is orthogonal to the line.*

15. $\mathbf{u} = (-1,6);\ \mathbf{v} = \left(\frac{3}{5}, \frac{4}{5}\right)$

16. $\mathbf{u} = (2,3);\ \mathbf{v} = \left(\frac{5}{13}, \frac{12}{13}\right)$

17. $\mathbf{u} = (2,3);\ \mathbf{v} = (1,1)$

18. $\mathbf{u} = (3,-1);\ \mathbf{v} = (3,4)$

In Exercises 19–22, let R^3 have the Euclidean inner product.

a. *Find the orthogonal projection of $\mathbf{u}$ onto the plane spanned by the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$.*

b. *Find the component of $\mathbf{u}$ orthogonal to the plane spanned by the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, and confirm that this component is orthogonal to the plane.*

19. $\mathbf{u} = (4,2,1);\ \mathbf{v}_1 = \left(\frac{1}{3}, \frac{2}{3}, -\frac{2}{3}\right),\ \mathbf{v}_2 = \left(\frac{2}{3}, \frac{1}{3}, \frac{2}{3}\right)$

20. $\mathbf{u} = (3,-1,2);\ \mathbf{v}_1 = \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right),\ \mathbf{v}_2 = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$

21. $\mathbf{u} = (1,0,3);\ \mathbf{v}_1 = (1,-2,1),\ \mathbf{v}_2 = (2,1,0)$

22. $\mathbf{u} = (1,0,2);\ \mathbf{v}_1 = (3,1,2),\ \mathbf{v}_2 = (-1,1,1)$

In Exercises 23–24, the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal with respect to the Euclidean inner product on R^4. Find the orthogonal projection of $\mathbf{b} = (1,2,0,-2)$ on the subspace W spanned by these vectors.

23. $\mathbf{v}_1 = (1,1,1,1),\ \mathbf{v}_2 = (1,1,-1,-1)$

24. $\mathbf{v}_1 = (0,1,-4,-1),\ \mathbf{v}_2 = (3,5,1,1)$

In Exercises 25–26, the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are orthonormal with respect to the Euclidean inner product on R^4. Find the orthogonal projection of $\mathbf{b} = (1,2,0,-1)$ onto the subspace W spanned by these vectors.

25. $\mathbf{v}_1 = \left(0, \frac{1}{\sqrt{18}}, -\frac{4}{\sqrt{18}}, -\frac{1}{\sqrt{18}}\right)$, $\mathbf{v}_2 = \left(\frac{1}{2}, \frac{5}{6}, \frac{1}{6}, \frac{1}{6}\right)$,
$\mathbf{v}_3 = \left(\frac{1}{\sqrt{18}}, 0, \frac{1}{\sqrt{18}}, -\frac{4}{\sqrt{18}}\right)$

26. $\mathbf{v}_1 = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$, $\mathbf{v}_2 = \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right)$, $\mathbf{v}_3 = \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)$

In Exercises 27–28, let R^2 have the Euclidean inner product and use the Gram–Schmidt process to transform the basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ into an orthonormal basis. Draw both sets of basis vectors in the xy-plane.

27. $\mathbf{u}_1 = (1, -3)$, $\mathbf{u}_2 = (2, 2)$ **28.** $\mathbf{u}_1 = (1, 0)$, $\mathbf{u}_2 = (3, -5)$

In Exercises 29–30, let R^3 have the Euclidean inner product and use the Gram–Schmidt process to transform the basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ into an orthonormal basis.

29. $\mathbf{u}_1 = (1, 1, 1)$, $\mathbf{u}_2 = (-1, 1, 0)$, $\mathbf{u}_3 = (1, 2, 1)$

30. $\mathbf{u}_1 = (1, 0, 0)$, $\mathbf{u}_2 = (3, 7, -2)$, $\mathbf{u}_3 = (0, 4, 1)$

31. Let R^4 have the Euclidean inner product. Use the Gram–Schmidt process to transform the basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ into an orthonormal basis.

$$\mathbf{u}_1 = (0, 2, 1, 0), \quad \mathbf{u}_2 = (1, -1, 0, 0),$$
$$\mathbf{u}_3 = (1, 2, 0, -1), \quad \mathbf{u}_4 = (1, 0, 0, 1)$$

32. Let R^3 have the Euclidean inner product. Find an orthonormal basis for the subspace spanned by $(0, 1, 2)$, $(-1, 0, 1)$, $(-1, 1, 3)$.

33. Let $\mathbf{b}$ and W be as in Exercise 23. Find vectors $\mathbf{w}_1$ in W and $\mathbf{w}_2$ in $W^\perp$ such that $\mathbf{b} = \mathbf{w}_1 + \mathbf{w}_2$.

34. Let $\mathbf{b}$ and W be as in Exercise 25. Find vectors $\mathbf{w}_1$ in W and $\mathbf{w}_2$ in $W^\perp$ such that $\mathbf{b} = \mathbf{w}_1 + \mathbf{w}_2$.

35. Let R^3 have the Euclidean inner product. The subspace of R^3 spanned by the vectors $\mathbf{u}_1 = (1, 1, 1)$ and $\mathbf{u}_2 = (2, 0, -1)$ is a plane passing through the origin. Express $\mathbf{w} = (1, 2, 3)$ in the form $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1$ lies in the plane and $\mathbf{w}_2$ is perpendicular to the plane.

36. Let R^4 have the Euclidean inner product. Express the vector $\mathbf{w} = (-1, 2, 6, 0)$ in the form $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1$ is in the space W that is spanned by $\mathbf{u}_1 = (-1, 0, 1, 2)$ and $\mathbf{u}_2 = (0, 1, 0, 1)$, and $\mathbf{w}_2$ is orthogonal to W.

37. Let R^3 have the inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + 2u_2 v_2 + 3u_3 v_3$$

Use the Gram–Schmidt process to transform $\mathbf{u}_1 = (1, 1, 1)$, $\mathbf{u}_2 = (1, 1, 0)$, $\mathbf{u}_3 = (1, 0, 0)$ into an orthonormal basis.

38. Verify that the set of vectors $\{(1, 0), (0, 1)\}$ is orthogonal with respect to the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1 v_1 + u_2 v_2$ on R^2; then convert it to an orthonormal set by normalizing the vectors.

39. Find vectors $\mathbf{x}$ and $\mathbf{y}$ in R^2 that are orthonormal with respect to the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1 v_1 + 2u_2 v_2$ but are not orthonormal with respect to the Euclidean inner product.

40. In Example 6 of Section 3.3 we found the orthogonal projection of the vector $\mathbf{x} = (1, 5)$ onto the line through the origin making an angle of $\pi/6$ radians with the positive x-axis. Solve that same problem using Theorem 6.3.4.

41. This exercise illustrates that the orthogonal projection resulting from Formula (12) in Theorem 6.3.4 does not depend on which orthogonal basis vectors are used.

a. Let R^3 have the Euclidean inner product, and let W be the subspace of R^3 spanned by the orthogonal vectors

$$\mathbf{v}_1 = (1, 0, 1) \quad \text{and} \quad \mathbf{v}_2 = (0, 1, 0)$$

Show that the orthogonal vectors

$$\mathbf{v}_1' = (1, 1, 1) \quad \text{and} \quad \mathbf{v}_2' = (1, -2, 1)$$

span the same subspace W.

b. Let $\mathbf{u} = (-3, 1, 7)$ and show that the same vector $\text{proj}_W \mathbf{u}$ results regardless of which of the bases in part (a) is used for its computation.

42. ***(Calculus required)*** Use Theorem 6.3.2(a) to express the following polynomials as linear combinations of the first three Legendre polynomials (see the Remark following Example 9).

a. $1 + x + 4x^2$ **b.** $2 - 7x^2$ **c.** $4 + 3x$

43. ***(Calculus required)*** Let P_2 have the inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_0^1 p(x)q(x)\, dx$$

Apply the Gram–Schmidt process to transform the standard basis $S = \{1, x, x^2\}$ into an orthonormal basis.

44. Find an orthogonal basis for the column space of the matrix

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

In Exercises 45–48, we obtained the column vectors of Q by applying the Gram–Schmidt process to the column vectors of A. Find a QR-decomposition of the matrix A.

45. $A = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}$, $Q = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$

46. $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 4 \end{bmatrix}$, $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix}$

47. $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 1 & 2 & 0 \end{bmatrix}$, $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \end{bmatrix}$

48. $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 3 & 1 \end{bmatrix}$, $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{\sqrt{2}}{2\sqrt{19}} & -\frac{3}{\sqrt{19}} \\ \frac{1}{\sqrt{2}} & -\frac{\sqrt{2}}{2\sqrt{19}} & \frac{3}{\sqrt{19}} \\ 0 & \frac{3\sqrt{2}}{\sqrt{19}} & \frac{1}{\sqrt{19}} \end{bmatrix}$

49. Find a QR-decomposition of the matrix

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

50. In the Remark following Example 8 we discussed two alternative ways to perform the calculations in the Gram–Schmidt process: normalizing each orthogonal basis vector as soon as it is calculated and scaling the orthogonal basis vectors at each step to eliminate fractions. Try these methods in Example 8.

Working with Proofs

51. Prove part (*a*) of Theorem 6.3.6.

52. In Step 3 of the proof of Theorem 6.3.5, it was stated that "the linear independence of $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ ensures that $\mathbf{v}_3 \neq \mathbf{0}$." Prove this statement.

53. Prove that the diagonal entries of R in Formula (16) are nonzero.

54. Show that matrix Q given in Example 10 satisfies the equation $QQ^T = I_3$, and prove that every $m \times n$ matrix Q with orthonormal column vectors has the property $QQ^T = I_m$.

55. a. Prove that if W is a subspace of a finite-dimensional vector space V, then the mapping $T: V \to W$ that is defined by $T(\mathbf{v}) = \text{proj}_W \mathbf{v}$ is a linear transformation.

b. What are the range and kernel of the transformation in part (a)?

True-False Exercises

TF. In parts **(a)–(f)** determine whether the statement is true or false, and justify your answer.

a. Every linearly independent set of vectors in an inner product space is orthogonal.

b. Every orthogonal set of vectors in an inner product space is linearly independent.

c. Every nontrivial subspace of R^3 has an orthonormal basis with respect to the Euclidean inner product.

d. Every nonzero finite-dimensional inner product space has an orthonormal basis.

e. $\text{proj}_W \mathbf{x}$ is orthogonal to every vector of W.

f. If A is an $n \times n$ matrix with a nonzero determinant, then A has a QR-decomposition.

Working with Technology

T1. a. Use the Gram–Schmidt process to find an orthonormal basis relative to the Euclidean inner product for the column space of

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

b. Use the method of Example 9 to find a QR-decomposition of A.

T2. Let P_4 have the evaluation inner product at the points $-2, -1, 0, 1, 2$. Find an orthogonal basis for P_4 relative to this inner product by applying the Gram–Schmidt process to the vectors

$$\mathbf{p}_0 = 1, \quad \mathbf{p}_1 = x, \quad \mathbf{p}_2 = x^2, \quad \mathbf{p}_3 = x^3, \quad \mathbf{p}_4 = x^4$$

6.4 Best Approximation; Least Squares

There are many applications in which some linear system $A\mathbf{x} = \mathbf{b}$ of m equations in n unknowns should be consistent on physical grounds but fails to be so because of measurement errors in the entries of A or $\mathbf{b}$. In such cases one looks for vectors that come as close as possible to being solutions in the sense that they minimize $\|\mathbf{b} - A\mathbf{x}\|$ with respect to the Euclidean inner product on R^m. In this section we will discuss methods for finding such minimizing vectors.

Least Squares Solutions of Linear Systems

Suppose that $A\mathbf{x} = \mathbf{b}$ is an *inconsistent* linear system of m equations in n unknowns in which we suspect the inconsistency to be caused by errors in the entries of A or $\mathbf{b}$. Since no exact solution is possible, we will look for a vector $\mathbf{x}$ that comes as "close as possible" to being a solution in the sense that it minimizes $\|\mathbf{b} - A\mathbf{x}\|$ with respect to the Euclidean inner product on R^m. You can think of $A\mathbf{x}$ as an approximation to $\mathbf{b}$ and $\|\mathbf{b} - A\mathbf{x}\|$ as the *error* in that approximation—the smaller the error, the better the approximation. This leads to the following problem.

Least Squares Problem Given a linear system $A\mathbf{x} = \mathbf{b}$ of m equations in n unknowns, find a vector $\mathbf{x}$ in R^n that minimizes $\|\mathbf{b} - A\mathbf{x}\|$ with respect to the Euclidean inner product on R^m. We call such a vector, if it exists, a ***least squares solution*** of the equation $A\mathbf{x} = \mathbf{b}$, we call $\mathbf{b} - A\mathbf{x}$ the ***least squares error vector***, and we call $\|\mathbf{b} - A\mathbf{x}\|$ the ***least squares error***.

If a linear system is consistent, then its exact solutions are the same as its least squares solutions, in which case the least squares error is zero.

To explain the terminology in this problem, suppose that the column form of $\mathbf{b} - A\mathbf{x}$ is

$$\mathbf{b} - A\mathbf{x} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}$$

The term "least squares solution" results from the fact that minimizing $\|\mathbf{b} - A\mathbf{x}\|$ also has the effect of minimizing

$$\|\mathbf{b} - A\mathbf{x}\|^2 = e_1^2 + e_2^2 + \cdots + e_m^2$$

What is important to keep in mind about the least squares problem is that for every vector $\mathbf{x}$ in R^n, the product $A\mathbf{x}$ is in the column space of A because it is a linear combination of the column vectors of A. That being the case, to find a least squares solution of $A\mathbf{x} = \mathbf{b}$ is equivalent to finding a vector $A\hat{\mathbf{x}}$ in the column space of A that is closest to $\mathbf{b}$ in the sense that it minimizes the length of the vector $\mathbf{b} - A\mathbf{x}$. This is illustrated in **Figure 6.4.1*a***, which also suggests that $A\hat{\mathbf{x}}$ is the orthogonal projection of $\mathbf{b}$ on the column space of A, that is, $A\hat{\mathbf{x}} = \text{proj}_{\text{col}(A)}\mathbf{b}$ (**Figure 6.4.1*b***). The next theorem will confirm this conjecture.

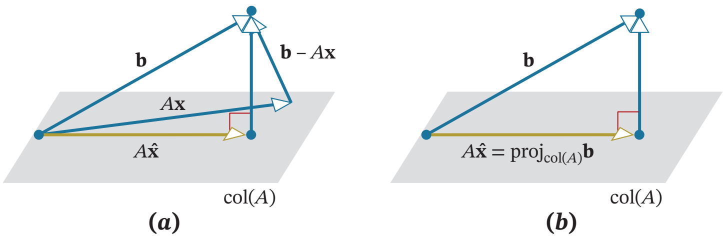


FIGURE 6.4.1

Theorem 6.4.1

Best Approximation Theorem

If W is a finite-dimensional subspace of an inner product space V, and if $\mathbf{b}$ is a vector in V, then $\text{proj}_W \mathbf{b}$ is the ***best approximation*** to $\mathbf{b}$ from W in the sense that

$$\|\mathbf{b} - \text{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\|$$

for every vector $\mathbf{w}$ in W that is different from $\text{proj}_W \mathbf{b}$.

Proof For every vector $\mathbf{w}$ in W, we can write

$$\mathbf{b} - \mathbf{w} = (\mathbf{b} - \text{proj}_W \mathbf{b}) + (\text{proj}_W \mathbf{b} - \mathbf{w}) \tag{1}$$

But $\text{proj}_W \mathbf{b} - \mathbf{w}$, being a difference of vectors in W, is itself in W; and since $\mathbf{b} - \text{proj}_W \mathbf{b}$ is orthogonal to W, the two terms on the right side of (1) are orthogonal. Thus, it follows from the Theorem of Pythagoras (Theorem 6.2.3) that

$$\|\mathbf{b} - \mathbf{w}\|^2 = \|\mathbf{b} - \text{proj}_W \mathbf{b}\|^2 + \|\text{proj}_W \mathbf{b} - \mathbf{w}\|^2$$

If $\mathbf{w} \neq \text{proj}_W \mathbf{b}$, it follows that the second term in this sum is positive, and hence that

$$\|\mathbf{b} - \text{proj}_W \mathbf{b}\|^2 < \|\mathbf{b} - \mathbf{w}\|^2$$

Taking square roots and using the fact that norms are nonnegative, it follows that

$$\|\mathbf{b} - \text{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\| \quad \blacksquare$$

It follows from Theorem 6.4.1 that if $V = R^n$ and $W = \text{col}(A)$, then the best approximation to $\mathbf{b}$ from $\text{col}(A)$ is $\text{proj}_{\text{col}(A)}\mathbf{b}$. But every vector in the column space of A is expressible in the form $A\mathbf{x}$ for some vector $\mathbf{x}$, so there is at least one vector $\hat{\mathbf{x}}$ in $\text{col}(A)$ for which $A\hat{\mathbf{x}} = \text{proj}_{\text{col}(A)}\mathbf{b}$. Each such vector is a least squares solution of $A\mathbf{x} = \mathbf{b}$, which shows that least squares solutions are not unique. Note, however, that although there may be more than one least squares solution of $A\mathbf{x} = \mathbf{b}$, each such solution $\hat{\mathbf{x}}$ has the same error vector $\mathbf{b} - A\hat{\mathbf{x}}$.

Finding Least Squares Solutions

One way to find a least squares solution of $A\mathbf{x} = \mathbf{b}$ is to calculate the orthogonal projection $\text{proj}_W \mathbf{b}$ on the column space W of A and then solve the equation

$$A\mathbf{x} = \text{proj}_W \mathbf{b} \tag{2}$$

However, we can avoid calculating the projection by rewriting (2) as

$$\mathbf{b} - A\mathbf{x} = \mathbf{b} - \text{proj}_W \mathbf{b}$$

and then multiplying both sides of this equation by A^T to obtain

$$A^T(\mathbf{b} - A\mathbf{x}) = A^T(\mathbf{b} - \text{proj}_W \mathbf{b}) \tag{3}$$

Since $\mathbf{b} - \text{proj}_W \mathbf{b}$ is the component of $\mathbf{b}$ that is orthogonal to the column space of A, it follows from Theorem 4.9.7(b) that this vector lies in the null space of A^T, and hence that

$$A^T(\mathbf{b} - \text{proj}_W \mathbf{b}) = \mathbf{0}$$

Thus, (3) simplifies to

$$A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$$

which we can rewrite as

$$A^TA\mathbf{x} = A^T\mathbf{b} \tag{4}$$

This is called the ***normal equation*** associated with $A\mathbf{x} = \mathbf{b}$. When viewed as a linear system, the individual equations are called the ***normal equations*** associated with $A\mathbf{x} = \mathbf{b}$.

In summary, we have established the following result.

Theorem 6.4.2

For every linear system $A\mathbf{x} = \mathbf{b}$, the associated normal system

$$A^TA\mathbf{x} = A^T\mathbf{b} \tag{5}$$

is consistent, and all solutions of (5) are least squares solutions of $A\mathbf{x} = \mathbf{b}$. Moreover, if $\mathbf{x}$ is any least squares solution, and W is the column space of A, then

$$A\mathbf{x} = \text{proj}_W \mathbf{b} \tag{6}$$

EXAMPLE 1 | Unique Least Squares Solution

Find a least squares solution, the least squares error vector, and the least squares error of the linear system

$$\begin{aligned} x_1 - x_2 &= 4 \\ 3x_1 + 2x_2 &= 1 \\ -2x_1 + 4x_2 &= 3 \end{aligned}$$

Solution It will be convenient to express the system in the matrix form $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} \tag{7}$$

It follows that

$$A^TA = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix} \tag{8}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 10 \end{bmatrix}$$

so the normal system $A^TA\mathbf{x} = A^T\mathbf{b}$ is

$$\begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 10 \end{bmatrix}$$

Solving this system yields a unique least squares solution, namely,

$$x_1 = \tfrac{17}{95}, \quad x_2 = \tfrac{143}{285}$$

The least squares error vector is

$$\mathbf{b} - A\mathbf{x} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} \frac{17}{95} \\ \frac{143}{285} \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} -\frac{92}{285} \\ \frac{439}{285} \\ \frac{94}{57} \end{bmatrix} = \begin{bmatrix} \frac{1232}{285} \\ -\frac{154}{285} \\ \frac{77}{57} \end{bmatrix}$$

and the least squares error is

$$\|\mathbf{b} - A\mathbf{x}\| \approx 4.556$$

The computations in the next example are a little tedious for hand computation, so in absence of a calculating utility you may want to just read through it for its ideas and logical flow.

EXAMPLE 2 | Infinitely Many Least Squares Solutions

Find a least squares solutions, the least squares error vector, and the least squares error of the linear system

$$\begin{aligned} 3x_1 + 2x_2 - x_3 &= 2 \\ x_1 - 4x_2 + 3x_3 &= -2 \\ x_1 + 10x_2 - 7x_3 &= 1 \end{aligned}$$

Solution The matrix form of the system is $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & -4 & 3 \\ 1 & 10 & -7 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

It follows that

$$A^TA = \begin{bmatrix} 11 & 12 & -7 \\ 12 & 120 & -84 \\ -7 & -84 & 59 \end{bmatrix} \quad \text{and} \quad A^T\mathbf{b} = \begin{bmatrix} 5 \\ 22 \\ -15 \end{bmatrix}$$

so the augmented matrix for the normal system $A^TA\mathbf{x} = A^T\mathbf{b}$ is

$$\left[\begin{array}{rrr|r} 11 & 12 & -7 & 5 \\ 12 & 120 & -84 & 22 \\ -7 & -84 & 59 & -15 \end{array}\right]$$

The reduced row echelon form of this matrix is

$$\left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{7} & \frac{2}{7} \\ 0 & 1 & -\frac{5}{7} & \frac{13}{84} \\ 0 & 0 & 0 & 0 \end{array}\right]$$

from which it follows that there are infinitely many least squares solutions, and that they are given by the parametric equations

$$\begin{aligned} x_1 &= \tfrac{2}{7} - \tfrac{1}{7}t \\ x_2 &= \tfrac{13}{84} + \tfrac{5}{7}t \\ x_3 &= t \end{aligned}$$

As a check, let us verify that all least squares solutions produce the same least squares error vector and the same least squares error. To see that this is so, we first compute

$$\mathbf{b} - A\mathbf{x} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 & 2 & -1 \\ 1 & -4 & 3 \\ 1 & 10 & -7 \end{bmatrix} \begin{bmatrix} \frac{2}{7} - \frac{1}{7}t \\ \frac{13}{84} + \frac{5}{7}t \\ t \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{7}{6} \\ -\frac{1}{3} \\ \frac{11}{6} \end{bmatrix} = \begin{bmatrix} \frac{5}{6} \\ -\frac{5}{3} \\ -\frac{5}{6} \end{bmatrix}$$

Since $\mathbf{b} - A\mathbf{x}$ does not depend on t, all least squares solutions produce the same error vector, namely

$$\|\mathbf{b} - A\mathbf{x}\| = \sqrt{\left(\tfrac{5}{6}\right)^2 + \left(-\tfrac{5}{3}\right)^2 + \left(-\tfrac{5}{6}\right)^2} = \tfrac{5}{6}\sqrt{6}$$

Conditions for Uniqueness of Least Squares Solutions

We know from Theorem 6.4.2 that the system $A^TA\mathbf{x} = A^T\mathbf{b}$ of normal equations for $A\mathbf{x} = \mathbf{b}$ is consistent. Thus, it follows from Theorem 1.6.1 that every linear system $A\mathbf{x} = \mathbf{b}$ has either one least squares solution (as in Example 1) or infinitely many least squares solutions (as in Example 2). Since A^TA is a square matrix, uniqueness occurs if A^TA is invertible; otherwise there are infinitely many least squares solutions. The following theorem provides a test for invertibility of A^TA using column vectors of A.

Theorem 6.4.3

If A is an $m \times n$ matrix, then the following are equivalent.

(*a*) The column vectors of A are linearly independent.

(*b*) A^TA is invertible.

Proof We will prove that $(a) \Rightarrow (b)$ and leave the proof that $(b) \Rightarrow (a)$ as an exercise.

$\boldsymbol{(a) \Rightarrow (b)}$ Assume that the column vectors of A are linearly independent. The matrix A^TA has size $n \times n$, so we can prove that this matrix is invertible by showing that the linear system $A^TA\mathbf{x} = \mathbf{0}$ has only the trivial solution. But if $\mathbf{x}$ is any solution of this system, then $A\mathbf{x}$ is in the null space of A^T and also in the column space of A. By Theorem 4.9.7(b) these spaces are orthogonal complements, so part (b) of Theorem 6.2.4 implies that $A\mathbf{x} = \mathbf{0}$. But A is assumed to have linearly independent column vectors, so it follows from parts (b) and (h) of Theorem 5.1.5 that $\mathbf{x} = \mathbf{0}$. ■

The next theorem, which follows directly from Theorems 6.4.2 and 6.4.3, gives an explicit formula for the least squares solution of a linear system in which the coefficient matrix has linearly independent column vectors.

Theorem 6.4.4

If A is an $m \times n$ matrix with linearly independent column vectors, then for every $m \times 1$ matrix $\mathbf{b}$, the linear system $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution. This solution is given by

$$\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b} \tag{9}$$

Moreover, if W is the column space of A, then

$$A\mathbf{x} = A(A^TA)^{-1}A^T\mathbf{b} = \text{proj}_W \mathbf{b} \tag{10}$$

EXAMPLE 3 | A Formula Solution to Example 1

Use Formula (9) and the matrices in Formulas (7) and (8) to find the least squares solution of the linear system in Example 1.

Solution We leave it for you to verify that

$$\begin{aligned}
\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b} &= \begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} \\
&= \frac{1}{285} \begin{bmatrix} 21 & 3 \\ 3 & 14 \end{bmatrix} \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{17}{95} \\ \frac{143}{285} \end{bmatrix}
\end{aligned}$$

which agrees with the result obtained in Example 1.

It follows from Formula (10) that the standard matrix for the orthogonal projection on the column space of a matrix A is

$$P = A(A^TA)^{-1}A^T \tag{11}$$

We will use this result in the next example.

EXAMPLE 4 | Orthogonal Projection on a Column Space

We showed in Formula (12) of Section 3.3 that the standard matrix for the orthogonal projection onto the line W through the origin of R^2 that makes an angle θ with the positive x-axis is

$$P_\theta = \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix}$$

Derive this result using Formula (11).

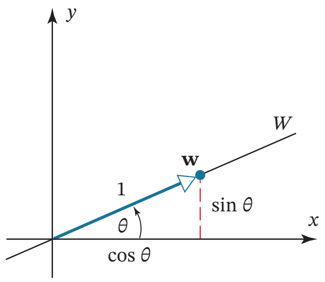


FIGURE 6.4.2

Solution To apply Formula (11) we must find a matrix A for which the line W is the column space. Since the line is one-dimensional and consists of all scalar multiples of the vector $\mathbf{w} = (\cos\theta, \sin\theta)$ (see Figure 6.4.2), we can take A to be

$$A = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

Since A^TA is the 1×1 identity matrix (verify), it follows that

$$A(A^TA)^{-1}A^T = AA^T = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} [\cos\theta \quad \sin\theta]$$

$$= \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix} = P_\theta$$

More on the Equivalence Theorem

As our next result we will add one additional part to Theorem 5.1.5.

Theorem 6.4.5

Equivalent Statements

If A is an $n \times n$ matrix in which there are no duplicate rows and no duplicate columns, then the following statements are equivalent.

(*a*) A is invertible.
(*b*) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
(*c*) The reduced row echelon form of A is I_n.
(*d*) A is expressible as a product of elementary matrices.
(*e*) $A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.
(*f*) $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.
(*g*) $\det(A) \neq 0$.
(*h*) The column vectors of A are linearly independent.
(*i*) The row vectors of A are linearly independent.
(*j*) The column vectors of A span R^n.
(*k*) The row vectors of A span R^n.
(*l*) The column vectors of A form a basis for R^n.
(*m*) The row vectors of A form a basis for R^n.
(*n*) A has rank n.
(*o*) A has nullity 0.
(*p*) The orthogonal complement of the null space of A is R^n.
(*q*) The orthogonal complement of the row space of A is $\{\mathbf{0}\}$.
(*r*) $\lambda = 0$ is not an eigenvalue of A.
(*s*) A^TA is invertible.

The proof of part (*s*) follows from part (*h*) of this theorem and Theorem 6.4.3 applied to square matrices.

OPTIONAL: Another View of Least Squares

Recall from Theorem 4.9.7 that the null space and row space of an $m \times n$ matrix A are orthogonal complements, as are the null space of A^T and the column space of A. Thus, given a linear system $A\mathbf{x} = \mathbf{b}$ in which A is an $m \times n$ matrix, Projection Theorem 6.3.3 tells us that the vectors $\mathbf{x}$ and $\mathbf{b}$ can each be decomposed into sums of orthogonal terms as

$$\mathbf{x} = \mathbf{x}_{\text{row}(A)} + \mathbf{x}_{\text{null}(A)} \quad \text{and} \quad \mathbf{b} = \mathbf{b}_{\text{null}(A^T)} + \mathbf{b}_{\text{col}(A)}$$

where $\mathbf{x}_{\text{row}(A)}$ and $\mathbf{x}_{\text{null}(A)}$ are the orthogonal projections of $\mathbf{x}$ on the row space of A and the null space of A, and the vectors $\mathbf{b}_{\text{null}(A^T)}$ and $\mathbf{b}_{\text{col}(A)}$ are the orthogonal projections of $\mathbf{b}$ on the null space of A^T and the column space of A.

In **Figure 6.4.3** we have represented the fundamental spaces of A by perpendicular lines in R^n and R^m on which we indicated the orthogonal projections of $\mathbf{x}$ and $\mathbf{b}$. (This, of course, is only pictorial since the fundamental spaces need not be one-dimensional.) The figure shows $A\mathbf{x}$ as a point in the column space of A and conveys that $\mathbf{b}_{\text{col}(A)}$ is the point in col(A) that is closest to $\mathbf{b}$. In the case where $A\mathbf{x} = \mathbf{b}$ is consistent, the vector $\mathbf{b}$ is in the column space of A, and the points $A\mathbf{x}$, $\mathbf{b}$, and $\mathbf{b}_{\text{col}(A)}$ coincide. The diagram indicates that multiplication by A maps $\mathbf{x}_{\text{row}(A)}$ into $A\mathbf{x}$. Explain why this is so.

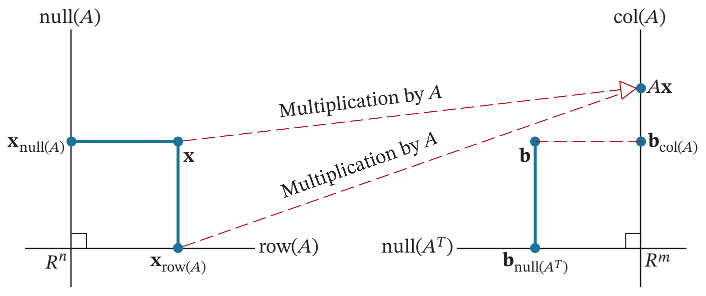


FIGURE 6.4.3

OPTIONAL: The Role of *QR*-Decomposition in Least Squares Problems

Formulas (9) and (10) have theoretical use but are not well suited for numerical computation. In practice, least squares solutions of $A\mathbf{x} = \mathbf{b}$ are typically found by using some variation of Gaussian elimination to solve the normal equations or by using QR-decomposition and the following theorem.

Theorem 6.4.6

If A is an $m \times n$ matrix with linearly independent column vectors, and if $A = QR$ is a QR-decomposition of A (see Theorem 6.3.7), then for each $\mathbf{b}$ in R^m the system $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution given by

$$\mathbf{x} = R^{-1}Q^T\mathbf{b} \tag{12}$$

A proof of this theorem and a discussion of its use can be found in many books on numerical methods of linear algebra. However, you can obtain Formula (12) by making the substitution $A = QR$ in (9) and using the fact that $Q^TQ = I$ to obtain

$$\begin{aligned} \mathbf{x} &= \big((QR)^T(QR)\big)^{-1}(QR)^T\mathbf{b} \\ &= (R^TQ^TQR)^{-1}(QR)^T\mathbf{b} \\ &= R^{-1}(R^T)^{-1}R^TQ^T\mathbf{b} \\ &= R^{-1}Q^T\mathbf{b} \end{aligned}$$

Exercise Set 6.4

In Exercises 1–2, find the associated normal equation.

1. $\begin{bmatrix} 1 & -1 \\ 2 & 3 \\ 4 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$

2. $\begin{bmatrix} 2 & -1 & 0 \\ 3 & 1 & 2 \\ -1 & 4 & 5 \\ 1 & 2 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$

In Exercises 3–6, find the least squares solution of the equation $A\mathbf{x} = \mathbf{b}$.

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

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

5. $A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & -2 \\ 1 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 9 \\ 3 \end{bmatrix}$

6. $A = \begin{bmatrix} 2 & 0 & -1 \\ 1 & -2 & 2 \\ 2 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 0 \\ 6 \\ 0 \\ 6 \end{bmatrix}$

In Exercises 7–10, find the least squares error vector and least squares error of the stated equation. Verify that the least squares error vector is orthogonal to the column space of A.

7. The equation in Exercise 3.

8. The equation in Exercise 4.

9. The equation in Exercise 5.

10. The equation in Exercise 6.

In Exercises 11–14, find parametric equations for all least squares solutions of $A\mathbf{x} = \mathbf{b}$, and confirm that all of the solutions have the same error vector.

11. $A = \begin{bmatrix} 2 & 1 \\ 4 & 2 \\ -2 & -1 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$

12. $A = \begin{bmatrix} 1 & 3 \\ -2 & -6 \\ 3 & 9 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

13. $A = \begin{bmatrix} -1 & 3 & 2 \\ 2 & 1 & 3 \\ 0 & 1 & 1 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 7 \\ 0 \\ -7 \end{bmatrix}$

14. $A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & -4 & 3 \\ 1 & 10 & -7 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$

In Exercises 15–16, use Theorem 6.4.2 to find the orthogonal projection of $\mathbf{b}$ on the column space of A, and check your result using Theorem 6.4.4.

15. $A = \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix}$

16. $A = \begin{bmatrix} 5 & 1 \\ 1 & 3 \\ 4 & -2 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} -4 \\ 2 \\ 3 \end{bmatrix}$

17. Find the orthogonal projection of $\mathbf{u}$ on the subspace of R^3 spanned by the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

$$\mathbf{u} = (1, -6, 1);\ \ \mathbf{v}_1 = (-1, 2, 1),\ \ \mathbf{v}_2 = (2, 2, 4)$$

18. Find the orthogonal projection of $\mathbf{u}$ on the subspace of R^4 spanned by the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$.

$$\mathbf{u} = (6, 3, 9, 6);\ \ \mathbf{v}_1 = (2, 1, 1, 1),\ \ \mathbf{v}_2 = (1, 0, 1, 1),$$
$$\mathbf{v}_3 = (-2, -1, 0, -1)$$

In Exercises 19–20, use the method of Example 3 to find the standard matrix for the orthogonal projection on the stated subspace of R^2. Compare your result to that in Table 3 of Section 1.8.

19. the x-axis

20. the y-axis

In Exercises 21–22, use the method of Example 3 to find the standard matrix for the orthogonal projection on the stated subspace of R^3. Compare your result to that in Table 4 of Section 1.8.

21. the xz-plane

22. the yz-plane

In Exercises 23–24, a QR-factorization of A is given. Use it to find the least squares solution of $A\mathbf{x} = \mathbf{b}$.

23. $A = \begin{bmatrix} 3 & 1 \\ -4 & 1 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix}\begin{bmatrix} 5 & -\frac{1}{5} \\ 0 & \frac{7}{5} \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

24. $A = \begin{bmatrix} 3 & -6 \\ 4 & -8 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & 0 \\ \frac{4}{5} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 5 & -10 \\ 0 & 1 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} -1 \\ 7 \\ 2 \end{bmatrix}$

25. Let W be the plane with equation $5x - 3y + z = 0$.

a. Find a basis for W.

b. Find the standard matrix for the orthogonal projection onto W.

26. Let W be the line with parametric equations

$$x = 2t, \quad y = -t, \quad z = 4t$$

a. Find a basis for W.

b. Find the standard matrix for the orthogonal projection on W.

27. Find the orthogonal projection of $\mathbf{u} = (5, 6, 7, 2)$ on the solution space of the homogeneous linear system

$$\begin{aligned} x_1 + x_2 + x_3 \phantom{{}+x_4} &= 0 \\ 2x_2 + x_3 + x_4 &= 0 \end{aligned}$$

28. Show that if $\mathbf{w} = (a, b, c)$ is a nonzero vector, then the standard matrix for the orthogonal projection of R^3 onto the line span$\{\mathbf{w}\}$ is

$$P = \frac{1}{a^2 + b^2 + c^2}\begin{bmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{bmatrix}$$

29. Let A be an $m \times n$ matrix with linearly independent row vectors. Find a standard matrix for the orthogonal projection of R^n onto the row space of A.

Working with Proofs

30. Prove: If A has linearly independent column vectors, and if $A\mathbf{x} = \mathbf{b}$ is consistent, then the least squares solution of the equation $A\mathbf{x} = \mathbf{b}$ and the exact solution of $A\mathbf{x} = \mathbf{b}$ are the same.

31. Prove: If A has linearly independent column vectors, and if $\mathbf{b}$ is orthogonal to the column space of A, then the least squares solution of $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = \mathbf{0}$.

32. Prove the implication $(b) \Rightarrow (a)$ of Theorem 6.4.3.

True-False Exercises

TF. In parts **(a)–(h)** determine whether the statement is true or false, and justify your answer.

a. If A is an $m \times n$ matrix, then A^TA is a square matrix.

b. If A^TA is invertible, then A is invertible.

c. If A is invertible, then A^TA is invertible.

d. If $A\mathbf{x} = \mathbf{b}$ is a consistent linear system, then $A^TA\mathbf{x} = A^T\mathbf{b}$ is also consistent.

e. If $A\mathbf{x} = \mathbf{b}$ is inconsistent, then $A^TA\mathbf{x} = A^T\mathbf{b}$ is also inconsistent.

f. Every linear system has a least squares solution.

g. Every linear system has a unique least squares solution.

h. If A is an $m \times n$ matrix with linearly independent columns and $\mathbf{b}$ is in R^m, then $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution.

Working with Technology

T1. **a.** Use Theorem 6.4.4 to show that the following linear system has a unique least squares solution, and use the method of Example 1 to find it.

$$\begin{aligned} x_1 + x_2 + x_3 &= 1 \\ 4x_1 + 2x_2 + x_3 &= 10 \\ 9x_1 + 3x_2 + x_3 &= 9 \\ 16x_1 + 4x_2 + x_3 &= 16 \end{aligned}$$

b. Check your result in part (a) using Formula (9).

T2. Use your technology utility to perform the computations and confirm the results obtained in Example 2.

6.5 Mathematical Modeling Using Least Squares

In this section we will use results about orthogonal projections in inner product spaces to obtain a method for fitting a line or other polynomial curve to a set of experimentally determined points in the plane.

Fitting a Curve to Data

A common problem in experimental work is to find a mathematical relationship $y = f(x)$ between two variables x and y by "fitting" a curve to points in the plane corresponding to various experimentally determined values of x and y, say

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

On the basis of theoretical considerations or simply by observing the pattern of the points, the experimenter decides on the general form of the curve $y = f(x)$ to be fitted. This curve is called a ***mathematical model*** of the data. Although mathematical models can be based on functions of other forms, we will focus on ***polynomial models***. Some examples are (**Figure 6.5.1**):

(a) A straight line: $y = a + bx$

(b) A quadratic polynomial: $y = a + bx + cx^2$

(c) A cubic polynomial: $y = a + bx + cx^2 + dx^3$

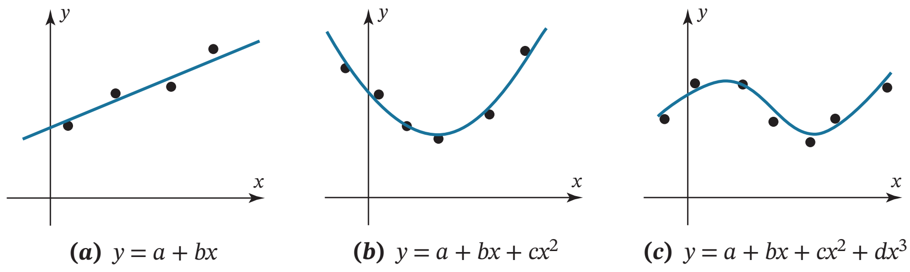


FIGURE 6.5.1

Least Squares Fit of a Straight Line

When data points are obtained experimentally, there is generally some measurement "error," making it impossible to find a curve of the desired form that passes through all the points. Thus, the idea is to choose the curve (by determining its coefficients) that "best fits" the data. We begin with the simplest case: fitting a straight line to data points.

Suppose we want to fit a straight line $y = a + bx$ to the experimentally determined points in which the x-coordinates are exact, but the y-coordinates may have errors, say

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

If the data points are collinear, the line will pass through all n points, and the unknown coefficients a and b will satisfy the equations

$$\begin{aligned} y_1 &= a + bx_1 \\ y_2 &= a + bx_2 \\ &\vdots \\ y_n &= a + bx_n \end{aligned} \tag{1}$$

We can write this system in matrix form as

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

or more compactly as

$$M\mathbf{v} = \mathbf{y} \tag{2}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} \tag{3}$$

If there are measurement errors in the data, then the data points will typically not lie on a line, and (1) will be inconsistent. In this case we look for a least squares approximation to the values of a and b by solving the normal system

$$M^T M\mathbf{v} = M^T\mathbf{y}$$

For simplicity, let us assume that the x-coordinates of the data points are not all the same, so M has linearly independent column vectors (why?) and the normal system has the unique solution

$$\mathbf{v}^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix} = (M^T M)^{-1} M^T\mathbf{y}$$

[see Formula (9) of Theorem 6.4.4]. The line $y = a^* + b^*x$ that results from this solution is called the ***regression line***. It follows from (2) and (3) that this line minimizes

$$\|\mathbf{y} - M\mathbf{v}\|^2 = [y_1 - (a + bx_1)]^2 + [y_2 - (a + bx_2)]^2 + \cdots + [y_n - (a + bx_n)]^2$$

The quantities

$$d_1 = |y_1 - (a + bx_1)|, \quad d_2 = |y_2 - (a + bx_2)|, \ldots, \quad d_n = |y_n - (a + bx_n)|$$

are called ***residuals***. Since the residual d_i is the distance between the data point (x_i, y_i) and the regression line (**Figure 6.5.2**), we can interpret its value as the "error" in y_i at the point x_i.

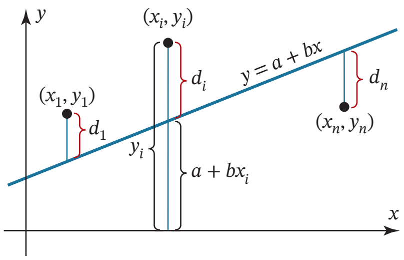


FIGURE 6.5.2 d_i measures the vertical error.

Since the regression line minimizes the sum of the squares of the data errors, it is commonly called the ***least squares line of best fit***.

Theorem 6.5.1

Uniqueness of the Regression Line

Let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be a set of two or more data points, not all lying on a vertical line, and let

$$M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{4}$$

Then there is a unique least squares straight line fit

$$y = a^* + b^*x \tag{5}$$

to the data points. Moreover,

$$\mathbf{v}^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix} \tag{6}$$

is given by the formula

$$\mathbf{v}^* = (M^TM)^{-1}M^T\mathbf{y} \tag{7}$$

which expresses the fact that $\mathbf{v} = \mathbf{v}^*$ is the unique solution of the normal equation

$$M^TM\mathbf{v} = M^T\mathbf{y} \tag{8}$$

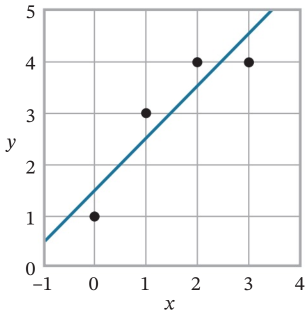


FIGURE 6.5.3

EXAMPLE 1 | Least Squares Straight Line Fit

Find the least squares straight line fit to the four points $(0,1)$, $(1,3)$, $(2,4)$, and $(3,4)$. (See **Figure 6.5.3**.)

Solution We have

$$M = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad M^TM = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}, \quad \text{and} \quad (M^TM)^{-1} = \frac{1}{10}\begin{bmatrix} 7 & -3 \\ -3 & 2 \end{bmatrix}$$

$$\mathbf{v}^* = (M^TM)^{-1}M^T\mathbf{y} = \frac{1}{10}\begin{bmatrix} 7 & -3 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1 \end{bmatrix}$$

so the desired line is $y = 1.5 + x$.

EXAMPLE 2 | Spring Constant

Hooke's law in physics states that the length x of a uniform spring is a linear function of the force y applied to it. If we express this relationship as $y = a + bx$, then the coefficient b is called the ***spring constant***. Suppose a particular unstretched spring has a measured length of 6.1 inches (i.e., $x = 6.1$ when $y = 0$). Suppose further that, as illustrated in **Figure 6.5.4**, various weights are attached to the end of the spring and that the following table of resulting spring lengths is recorded. Find the least squares straight line fit to the data and use it to approximate the spring constant.

Weight *y* (lb)	0	2	4	6
Length *x* (in.)	6.1	7.6	8.7	10.4

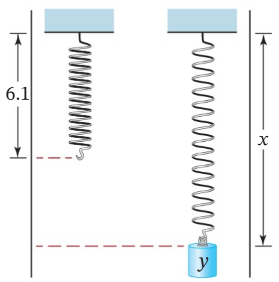


FIGURE 6.5.4

Solution The mathematical problem is to fit a line $y = a + bx$ to the four data points

$$(6.1, 0), \quad (7.6, 2), \quad (8.7, 4), \quad (10.4, 6)$$

For these data the matrices M and $\mathbf{y}$ in (4) are

$$M = \begin{bmatrix} 1 & 6.1 \\ 1 & 7.6 \\ 1 & 8.7 \\ 1 & 10.4 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 0 \\ 2 \\ 4 \\ 6 \end{bmatrix}$$

so

$$\mathbf{v}^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix} = (M^TM)^{-1}M^T\mathbf{y} \approx \begin{bmatrix} -8.6 \\ 1.4 \end{bmatrix}$$

where the numerical values have been rounded to one decimal place. Thus, the estimated value of the spring constant is $b^* \approx 1.4$ pounds/inch.

Least Squares Fit of a Polynomial

The technique described for fitting a straight line to data points can be generalized to fitting a polynomial of specified degree to data points. Let us attempt to fit a polynomial of fixed degree m

$$y = a_0 + a_1x + \cdots + a_mx^m \tag{9}$$

to n points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

Substituting these n values of x and y into (9) yields the n equations

$$\begin{aligned}
y_1 &= a_0 + a_1x_1 + \cdots + a_mx_1^m \\
y_2 &= a_0 + a_1x_2 + \cdots + a_mx_2^m \\
&\;\vdots \qquad \vdots \qquad\quad \vdots \qquad\qquad \vdots \\
y_n &= a_0 + a_1x_n + \cdots + a_mx_n^m
\end{aligned}$$

or in matrix form,

$$\mathbf{y} = M\mathbf{v} \tag{10}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad M = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} \tag{11}$$

As before, the solutions of the normal equations

$$M^TM\mathbf{v} = M^T\mathbf{y}$$

determine the coefficients of the polynomial, and the vector $\mathbf{v}$ minimizes

$$\|\mathbf{y} - M\mathbf{v}\|$$

Conditions that guarantee the invertibility of M^TM are discussed in the exercises. If M^TM is invertible, then the normal equations have a unique solution $\mathbf{v} = \mathbf{v}^*$, which is given by

$$\mathbf{v}^* = (M^TM)^{-1}M^T\mathbf{y} \tag{12}$$

EXAMPLE 3 | Fitting a Quadratic Curve to Data

According to Newton's second law of motion, a body near the Earth's surface falls vertically in accordance with the equation

$$s = s_0 + v_0t + \tfrac{1}{2}gt^2 \tag{13}$$

where

$$\begin{aligned}
s &= \text{vertical displacement downward relative to some reference point} \\
s_0 &= \text{displacement from the reference point at time } t = 0 \\
v_0 &= \text{velocity at time } t = 0 \\
g &= \text{acceleration of gravity at the Earth's surface}
\end{aligned}$$

Suppose that a laboratory experiment is performed to approximate g by measuring the displacement s relative to a fixed reference point of a falling weight at various times. Use the experimental results shown in the following table to approximate g.

Time t (sec)	.1	.2	.3	.4	.5
Displacement s (ft)	−0.18	0.31	1.03	2.48	3.73

Solution For notational simplicity, let $a_0 = s_0$, $a_1 = v_0$, and $a_2 = \frac{1}{2}g$ in (13), so our mathematical problem is to fit a quadratic curve

$$s = a_0 + a_1t + a_2t^2 \tag{14}$$

to the five data points:

$$(.1, -0.18), \quad (.2, 0.31), \quad (.3, 1.03), \quad (.4, 2.48), \quad (.5, 3.73)$$

With the appropriate adjustments in notation, the matrices M and $\mathbf{y}$ in (11) are

$$M = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \\ 1 & t_4 & t_4^2 \\ 1 & t_5 & t_5^2 \end{bmatrix} = \begin{bmatrix} 1 & .1 & .01 \\ 1 & .2 & .04 \\ 1 & .3 & .09 \\ 1 & .4 & .16 \\ 1 & .5 & .25 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{bmatrix} = \begin{bmatrix} -0.18 \\ 0.31 \\ 1.03 \\ 2.48 \\ 3.73 \end{bmatrix}$$

Thus, from (12),

$$\mathbf{v}^* = \begin{bmatrix} a_0^* \\ a_1^* \\ a_2^* \end{bmatrix} = (M^TM)^{-1}M^T\mathbf{y} \approx \begin{bmatrix} -0.40 \\ 0.35 \\ 16.1 \end{bmatrix}$$

so the least squares quadratic fit is

$$s = -0.40 + 0.35t + 16.1t^2$$

From this equation we estimate that $\frac{1}{2}g = 16.1$ and hence that $g = 32.2$ ft/sec^2. Note that this equation also provides the following estimates of the initial displacement and velocity of the weight:

$$s_0 = a_0^* = -0.40 \text{ ft}$$
$$v_0 = a_1^* = 0.35 \text{ ft/sec}$$

In **Figure 6.5.5** we have plotted the data points and the approximating polynomial.

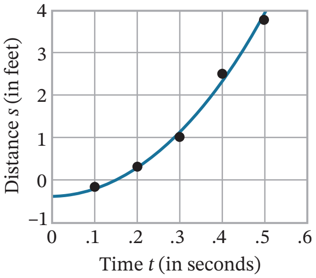
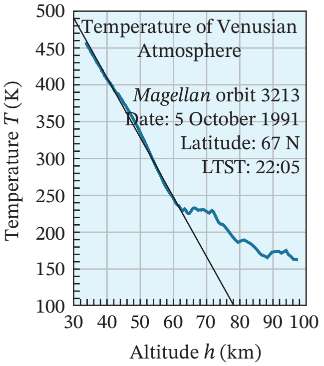


FIGURE 6.5.5

Historical Note



Source: NASA

On October 5, 1991 the *Magellan* spacecraft entered the atmosphere of Venus and transmitted the temperature T in kelvins (K) versus the altitude h in kilometers (km) until its signal was lost at an altitude of about 34 km. Discounting the initial erratic signal, the data strongly suggested a linear relationship, so a least squares straight line fit was used on the linear part of the data to obtain the equation

$$T = 737.5 - 8.125h$$

By setting $h = 0$ in this equation, the surface temperature of Venus was estimated at $T \approx 737.5$ K. The accuracy of this result has been confirmed by more recent flybys of Venus.

Exercise Set 6.5

In Exercises 1–2, find the least squares straight line fit

$$y = ax + b$$

to the data points, and show that the result is reasonable by graphing the fitted line and plotting the data in the same coordinate system.

1. $(0, 0), (1, 2), (2, 7)$

2. $(0, 1), (2, 0), (3, 1), (3, 2)$

In Exercises 3–4, find the least squares quadratic fit

$$y = a_0 + a_1x + a_2x^2$$

to the data points, and show that the result is reasonable by graphing the fitted curve and plotting the data in the same coordinate system.

3. $(2, 0), (3, -10), (5, -48), (6, -76)$

4. $(1,-2),\ (0,-1),\ (1,0),\ (2,4)$

5. Find a curve of the form $y = a + (b/x)$ that best fits the data points $(1,7), (3,3), (6,1)$ by making the substitution $X = 1/x$.

6. Find a curve of the form $y = a + b\sqrt{x}$ that best fits the data points $(3,1.5),\ (7,2.5),\ (10,3)$ by making the substitution $X = \sqrt{x}$. Show that the result is reasonable by graphing the fitted curve and plotting the data in the same coordinate system.

Working with Proofs

7. Prove that the matrix M in Equation (3) has linearly independent columns if and only if at least two of the numbers $x_1, x_2, \ldots, x_n$ are distinct.

8. Prove that the columns of the $n \times (m+1)$ matrix M in Equation (11) are linearly independent if $n > m$ and at least $m+1$ of the numbers $x_1, x_2, \ldots, x_n$ are distinct. [*Hint:* A nonzero polynomial of degree m has at most m distinct roots.]

9. Let M be the matrix in Equation (11). Using Exercise 8, show that a sufficient condition for the matrix M^TM to be invertible is that $n > m$ and that at least $m+1$ of the numbers $x_1, x_2, \ldots, x_n$ are distinct.

True-False Exercises

TF. In parts **(a)–(d)** determine whether the statement is true or false, and justify your answer.

a. Every set of data points has a unique least squares straight line fit.

b. If the data points $(x_1,y_1), (x_2,y_2), \ldots, (x_n,y_n)$ are not collinear, then (2) is an inconsistent system.

c. If the data points $(x_1,y_1), (x_2,y_2), \ldots, (x_n,y_n)$ do not lie on a vertical line, then the expression

$$|y_1 - (a+bx_1)|^2 + |y_2 - (a+bx_2)^2| + \cdots + |y_n - (a+bx_n)|^2$$

is minimized by taking a and b to be the coefficients in the least squares line $y = a + bx$ of best fit to the data.

d. If the data points $(x_1,y_1), (x_2,y_2), \ldots, (x_n,y_n)$ do not lie on a vertical line, then the expression

$$|y_1 - (a+bx_1)| + |y_2 - (a+bx_2)| + \cdots + |y_n - (a+bx_n)|$$

is minimized by taking a and b to be the coefficients in the least squares line $y = a + bx$ of best fit to the data.

Working with Technology

In Exercises T1–T7, find the normal system for the least squares cubic fit $y = a_0 + a_1x + a_2x^2 + a_3x^3$ to the data points. Solve the system and show that the result is reasonable by graphing the fitted curve and plotting the data in the same coordinate system.

T1. $(-1,-14),\ (0,-5),\ (1,-4),\ (2,1),\ (3,22)$

T2. $(0,-10),\ (1,-1),\ (2,0),\ (3,5),\ (4,26)$

T3. The owner of a rapidly expanding business finds that for the first five months of the year the sales (in thousands) are \$4.0, \$4.4, \$5.2, \$6.4, and \$8.0. The owner plots these figures on a graph and conjectures that for the rest of the year, the sales curve can be approximated by a quadratic polynomial. Find the least squares quadratic polynomial fit to the sales curve, and use it to project the sales for the twelfth month of the year.

T4. *Pathfinder* is an experimental, lightweight, remotely piloted, solar-powered aircraft that was used in a series of experiments by NASA to determine the feasibility of applying solar power for long-duration, high-altitude flights. In August 1997 *Pathfinder* recorded the data in the accompanying table relating altitude H and temperature T. Show that a linear model is reasonable by plotting the data, and then find the least squares line $H = H_0 + kT$ of best fit.

TABLE Ex-T4

Altitude H (thousands of feet)	15	20	25	30	35	40	45
Temperature T (°C)	4.5	−5.9	−16.1	−27.6	−39.8	−50.2	−62.9

Three important models in applications are

exponential models $(y = ae^{bx})$
power function models $(y = ax^b)$
logarithmic models $(y = a + b\ln x)$

where a and b are to be determined to fit experimental data as closely as possible. Exercises T5–T7 are concerned with a procedure, called ***linearization****, by which the data are transformed to a form in which a least squares straight line fit can be used to approximate the constants. Calculus is required for these exercises.*

T5. a. Show that making the substitution $Y = \ln y$ in the equation $y = ae^{bx}$ produces the equation $Y = bx + \ln a$ whose graph in the xY-plane is a line of slope b and Y-intercept $\ln a$.

b. Part (a) suggests that a curve of the form $y = ae^{bx}$ can be fitted to n data points (x_i, y_i) by letting $Y_i = \ln y_i$, then fitting a straight line to the transformed data points (x_i, Y_i) by least squares to find b and $\ln a$, and then computing a from $\ln a$. Use this method to fit an exponential model to the following data, and graph the curve and data in the same coordinate system.

x	0	1	2	3	4	5	6	7
y	3.9	5.3	7.2	9.6	12	17	23	31

T6. a. Show that making the substitutions

$$X = \ln x \quad \text{and} \quad Y = \ln y$$

in the equation $y = ax^b$ produces the equation

$$Y = bX + \ln a$$

whose graph in the XY-plane is a line of slope b and Y-intercept $\ln a$.

b. Part (a) suggests that a curve of the form $y = ax^b$ can be fitted to n data points (x_i, y_i) by letting $X_i = \ln x_i$ and $Y_i = \ln y_i$, then fitting a straight line to the transformed data points (X_i, Y_i) by least squares to find b and $\ln a$, and then computing a from $\ln a$. Use this method to fit a power function model to the following data, and graph the curve and data in the same coordinate system.

x	2	3	4	5	6	7	8	9
y	1.75	1.91	2.03	2.13	2.22	2.30	2.37	2.43

T7. a. Show that making the substitution $X = \ln x$ in the equation $y = a + b \ln x$ produces the equation $y = a + bX$ whose graph in the Xy-plane is a line of slope b and y-intercept a.

b. Part (a) suggests that a curve of the form $y = a + b \ln x$ can be fitted to n data points (x_i, y_i) by letting $X_i = \ln x_i$ and then fitting a straight line to the transformed data points (X_i, y_i) by least squares to find b and a. Use this method to fit a logarithmic model to the following data, and graph the curve and data in the same coordinate system.

x	2	3	4	5	6	7	8	9
y	4.07	5.30	6.21	6.79	7.32	7.91	8.23	8.51

6.6 Function Approximation; Fourier Series

In this section we will show how orthogonal projections can be used to approximate certain types of functions by simpler functions. The ideas explained here have important applications in engineering and science. Calculus is required for this section.

Best Approximations

All of the problems that we will study in this section will be special cases of the following general problem.

Approximation Problem Given a function f that is continuous on an interval $[a, b]$, find the "best possible approximation" to f using only functions from a specified subspace W of $C[a, b]$.

Here are some examples of such problems:

(a) Find the best possible approximation to e^x over the interval $[0, 1]$ by a polynomial of the form $a_0 + a_1x + a_2x^2$.

(b) Find the best possible approximation to $\sin \pi x$ over the interval $[-1, 1]$ by a function of the form $a_0 + a_1e^x + a_2e^{2x} + a_3e^{3x}$.

(c) Find the best possible approximation to x over the interval $[0, 2\pi]$ by a function of the form $a_0 + a_1 \sin x + a_2 \sin 2x + b_1 \cos x + b_2 \cos 2x$.

In the first example W is the subspace of $C[0, 1]$ spanned by $1, x$, and x^2; in the second example W is the subspace of $C[-1, 1]$ spanned by $1, e^x$, e^{2x}, and e^{3x}; and in the third example W is the subspace of $C[0, 2\pi]$ spanned by $1, \sin x, \sin 2x, \cos x$, and $\cos 2x$.

Measurements of Error

To solve approximation problems of the preceding types, we first need to make the phrase "best approximation over $[a, b]$" mathematically precise. To do this we will need some way of quantifying the error that results when one continuous function is approximated by another over an interval $[a, b]$. If we were to approximate $f(x)$ by $g(x)$, and if we were concerned only with the error in that approximation at a *single point* x_0, then it would be natural to define the error to be

$$\text{error} = |f(x_0) - g(x_0)|$$

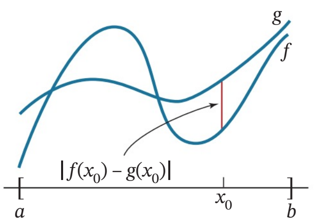


FIGURE 6.6.1 The deviation between f and g at x_0.

sometimes called the ***deviation*** between f and g at x_0 (**Figure 6.6.1**). However, we are not concerned simply with measuring the error at a single point but rather with measuring it over the *entire* interval $[a, b]$. The difficulty is that an approximation may have small deviations in one part of the interval and large deviations in another. One possible way of accounting for this is to integrate the deviation $|f(x) - g(x)|$ over the interval $[a, b]$ and define the error over the interval to be

$$\text{error} = \int_a^b |f(x) - g(x)|\, dx \tag{1}$$

Geometrically, (1) is the area between the graphs of $f(x)$ and $g(x)$ over the interval $[a, b]$ (**Figure 6.6.2**)— the greater the area, the greater the overall error.

Although (1) is natural and appealing geometrically, most mathematicians and scientists generally favor the following alternative measure of error, called the ***mean square error***.

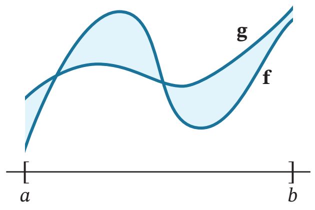


FIGURE 6.6.2 The area between the graphs of **f** and **g** over $[a, b]$ measures the error in approximating f by g over $[a, b]$.

$$\boxed{\text{mean square error} = \int_a^b [f(x) - g(x)]^2\, dx}$$

Although mean square error emphasizes larger deviations because of the squaring, it has the advantage of allowing us to bring to bear the theory of inner product spaces. To see how, suppose that **f** is a continuous function on $[a, b]$ that we want to approximate by a function **g** from a subspace W of $C[a, b]$, and suppose that $C[a, b]$ is given the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x)\, dx$$

It follows that

$$\|\mathbf{f} - \mathbf{g}\|^2 = \langle \mathbf{f} - \mathbf{g}, \mathbf{f} - \mathbf{g} \rangle = \int_a^b [f(x) - g(x)]^2\, dx = \text{mean square error}$$

so minimizing the mean square error is the same as minimizing $\|\mathbf{f} - \mathbf{g}\|^2$. Thus, the approximation problem posed informally at the beginning of this section can be restated more precisely as follows.

Least Squares Approximation

Least Squares Approximation Problem Let **f** be a function that is continuous on an interval $[a, b]$, let $C[a, b]$ have the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x)\, dx$$

and let W be a finite-dimensional subspace of $C[a, b]$. Find a function **g** in W that minimizes

$$\|\mathbf{f} - \mathbf{g}\|^2 = \int_a^b [f(x) - g(x)]^2\, dx$$

Since $\|\mathbf{f} - \mathbf{g}\|^2$ and $\|\mathbf{f} - \mathbf{g}\|$ are minimized by the same function **g**, this problem is equivalent to looking for a function **g** in W that is closest to **f**. But we know from Theorem 6.4.1 that $\mathbf{g} = \text{proj}_W \mathbf{f}$ is such a function (**Figure 6.6.3**).

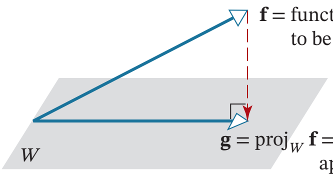


FIGURE 6.6.3

Thus, we have the following result.

Theorem 6.6.1

If $\mathbf{f}$ is a continuous function on $[a, b]$, and W is a finite-dimensional subspace of $C[a, b]$, then the function $\mathbf{g}$ in W that minimizes the mean square error

$$\int_a^b [f(x) - g(x)]^2\, dx$$

is $\mathbf{g} = \text{proj}_W \mathbf{f}$, where the orthogonal projection is relative to the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x)\, dx$$

The function $\mathbf{g} = \text{proj}_W \mathbf{f}$ is called the ***least squares approximation*** to $\mathbf{f}$ from W.

Fourier Series

A function of the form

$$\begin{aligned} T(x) = c_0 + c_1 \cos x + c_2 \cos 2x + \cdots + c_n \cos nx \\ + d_1 \sin x + d_2 \sin 2x + \cdots + d_n \sin nx \end{aligned} \tag{2}$$

is called a ***trigonometric polynomial***; if c_n and d_n are not both zero, then $T(x)$ is said to have ***order n***. For example,

$$T(x) = 2 + \cos x - 3\cos 2x + 7\sin 4x$$

is a trigonometric polynomial of order 4 with

$$c_0 = 2, \quad c_1 = 1, \quad c_2 = -3, \quad c_3 = 0, \quad c_4 = 0, \quad d_1 = 0, \quad d_2 = 0, \quad d_3 = 0, \quad d_4 = 7$$

It is evident from (2) that the trigonometric polynomials of order n or less are the various possible linear combinations of

$$1, \ \cos x, \ \cos 2x, \ldots, \ \cos nx, \qquad \sin x, \ \sin 2x, \ldots, \ \sin nx \tag{3}$$

It can be shown that these $2n + 1$ functions are linearly independent and thus form a basis for a $(2n + 1)$-dimensional subspace of $C[a, b]$.

Let us now consider the problem of finding the least squares approximation of a continuous function $f(x)$ over the interval $[0, 2\pi]$ by a trigonometric polynomial of order n or less. As noted above, the least squares approximation to $\mathbf{f}$ from W is the orthogonal projection of $\mathbf{f}$ on W. To find this orthogonal projection, we must find an orthonormal basis $\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_{2n}$ for W, after which we can compute the orthogonal projection on W from the formula

$$\text{proj}_W \mathbf{f} = \langle \mathbf{f}, \mathbf{g}_0 \rangle \mathbf{g}_0 + \langle \mathbf{f}, \mathbf{g}_1 \rangle \mathbf{g}_1 + \cdots + \langle \mathbf{f}, \mathbf{g}_{2n} \rangle \mathbf{g}_{2n} \tag{4}$$

[see Theorem 6.3.4(b)]. An orthonormal basis for W can be obtained by applying the Gram–Schmidt process to the basis vectors in (3) using the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_0^{2\pi} f(x)g(x)\, dx$$

This yields the orthonormal basis

$$\begin{aligned} \mathbf{g}_0 = \frac{1}{\sqrt{2\pi}}, \quad \mathbf{g}_1 = \frac{1}{\sqrt{\pi}} \cos x, \ldots, \quad \mathbf{g}_n = \frac{1}{\sqrt{\pi}} \cos nx, \\ \mathbf{g}_{n+1} = \frac{1}{\sqrt{\pi}} \sin x, \ldots, \quad \mathbf{g}_{2n} = \frac{1}{\sqrt{\pi}} \sin nx \end{aligned} \tag{5}$$

(see Exercise 6). If we introduce the notation

$$\begin{aligned} a_0 = \frac{2}{\sqrt{2\pi}} \langle \mathbf{f}, \mathbf{g}_0 \rangle, \quad a_1 = \frac{1}{\sqrt{\pi}} \langle \mathbf{f}, \mathbf{g}_1 \rangle, \ldots, \quad a_n = \frac{1}{\sqrt{\pi}} \langle \mathbf{f}, \mathbf{g}_n \rangle \\ b_1 = \frac{1}{\sqrt{\pi}} \langle \mathbf{f}, \mathbf{g}_{n+1} \rangle, \ldots, \quad b_n = \frac{1}{\sqrt{\pi}} \langle \mathbf{f}, \mathbf{g}_{2n} \rangle \end{aligned} \tag{6}$$

then on substituting (5) in (4), we obtain

$$\boxed{\operatorname{proj}_W \mathbf{f} = \frac{a_0}{2} + [a_1 \cos x + \cdots + a_n \cos nx] + [b_1 \sin x + \cdots + b_n \sin nx]} \tag{7}$$

where

$$\begin{aligned}
a_0 &= \frac{2}{\sqrt{2\pi}}\langle \mathbf{f}, \mathbf{g}_0\rangle = \frac{2}{\sqrt{2\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{2\pi}}\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\,dx \\
a_1 &= \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, \mathbf{g}_1\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\cos x\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos x\,dx \\
&\ \vdots \\
a_n &= \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, \mathbf{g}_n\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\cos nx\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos nx\,dx \\
b_1 &= \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, \mathbf{g}_{n+1}\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\sin x\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin x\,dx \\
&\ \vdots \\
b_n &= \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, \mathbf{g}_{2n}\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\sin nx\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin nx\,dx
\end{aligned}$$

In short,

$$\boxed{a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx, \quad b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,dx} \tag{8}$$

The numbers $a_0, a_1, \ldots, a_n, b_1, \ldots, b_n$ are called the ***Fourier coefficients*** of $\mathbf{f}$.

EXAMPLE 1 | Least Squares Approximations

Find the least squares approximation of $f(x) = x$ on $[0, 2\pi]$ by

(a) a trigonometric polynomial of order 2 or less.

(b) a trigonometric polynomial of order n or less.

Solution (a)

$$a_0 = \frac{1}{\pi}\int_0^{2\pi} f(x)\,dx = \frac{1}{\pi}\int_0^{2\pi} x\,dx = 2\pi \tag{9a}$$

For $k = 1, 2, \ldots,$ integration by parts yields (verify)

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx = \frac{1}{\pi}\int_0^{2\pi} x\cos kx\,dx = 0 \tag{9b}$$

$$b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,dx = \frac{1}{\pi}\int_0^{2\pi} x\sin kx\,dx = -\frac{2}{k} \tag{9c}$$

Thus, the least squares approximation to x on $[0, 2\pi]$ by a trigonometric polynomial of order 2 or less is

$$x \approx \frac{a_0}{2} + a_1 \cos x + a_2 \cos 2x + b_1 \sin x + b_2 \sin 2x$$

or, from (9a), (9b), and (9c),

$$x \approx \pi - 2\sin x - \sin 2x$$

Solution (*b*) The least squares approximation to x on $[0, 2\pi]$ by a trigonometric polynomial of order n or less is

$$x \approx \frac{a_0}{2} + [a_1 \cos x + \cdots + a_n \cos nx] + [b_1 \sin x + \cdots + b_n \sin nx]$$

or, from (9a), (9b), and (9c),

$$x \approx \pi - 2\left(\sin x + \frac{\sin 2x}{2} + \frac{\sin 3x}{3} + \cdots + \frac{\sin nx}{n}\right)$$

The graphs of $y = x$ and some of these approximations are shown in **Figure 6.6.4**.

It is natural to expect that the mean square error will diminish as the number of terms in the least squares approximation

$$f(x) \approx \frac{a_0}{2} + \sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx)$$

increase. It can be proved that for functions f in $C[0, 2\pi]$, the mean square error approaches zero as $n \to +\infty$; this is denoted by writing

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx)$$

The right side of this equation is called the ***Fourier series*** for f over the interval $[0, 2\pi]$. Such series are of major importance in engineering, science, and mathematics.

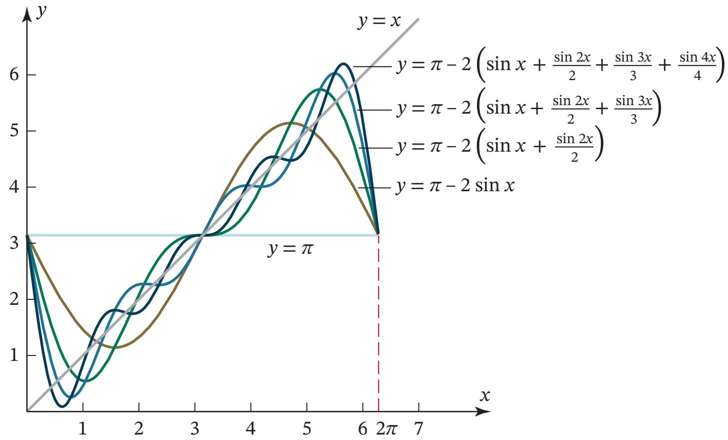


FIGURE 6.6.4

Historical Note

Jean Baptiste Fourier (1768–1830)

Fourier was a French mathematician and physicist who discovered the Fourier series and related ideas while working on problems of heat diffusion. This discovery was one of the most influential in the history of mathematics; it is the cornerstone of many fields of mathematical research and a basic tool in many branches of engineering. Fourier, a political activist during the French revolution, spent time in jail for his defense of many victims during the Reign of Terror. He later became a favorite of Napoleon who made him a baron.

[Image: Hulton Archive/Getty Images]

Exercise Set 6.6

1. Find the least squares approximation of $f(x) = 1 + x$ over the interval $[0, 2\pi]$ by
   a. a trigonometric polynomial of order 2 or less.
   b. a trigonometric polynomial of order n or less.

2. Find the least squares approximation of $f(x) = x^2$ over the interval $[0, 2\pi]$ by
   a. a trigonometric polynomial of order 3 or less.
   b. a trigonometric polynomial of order n or less.

3. a. Find the least squares approximation of x over the interval $[0, 1]$ by a function of the form $a + be^x$.
   b. Find the mean square error of the approximation.

4. a. Find the least squares approximation of e^x over the interval $[0, 1]$ by a polynomial of the form $a_0 + a_1x$.
   b. Find the mean square error of the approximation.

5. a. Find the least squares approximation of $\sin \pi x$ over the interval $[-1, 1]$ by a polynomial of the form $a_0 + a_1x + a_2x^2$.
   b. Find the mean square error of the approximation.

6. Use the Gram–Schmidt process to obtain the orthonormal basis (5) from the basis (3).

7. Carry out the integrations indicated in Formulas (9a), (9b), and (9c).

8. Find the Fourier series of $f(x) = \pi - x$ over the interval $[0, 2\pi]$.

9. Find the Fourier series of $f(x) = 1, 0 < x < \pi$ and $f(x) = 0, \pi \leq x \leq 2\pi$ over the interval $[0, 2\pi]$.

10. What is the Fourier series of $\sin(3x)$?

True-False Exercises

TF. In parts (a)–(e) determine whether the statement is true or false, and justify your answer.

a. If a function $\mathbf{f}$ in $C[a, b]$ is approximated by the function $\mathbf{g}$, then the mean square error is the same as the area between the graphs of $f(x)$ and $g(x)$ over the interval $[a, b]$.

b. Given a finite-dimensional subspace W of $C[a, b]$, the function $\mathbf{g} = \text{proj}_W \mathbf{f}$ minimizes the mean square error.

c. $\{1, \cos x, \sin x, \cos 2x, \sin 2x\}$ is an orthogonal subset of the vector space $C[0, 2\pi]$ with respect to the inner product $\langle \mathbf{f}, \mathbf{g} \rangle = \int_0^{2\pi} f(x)g(x)\,dx$.

d. $\{1, \cos x, \sin x, \cos 2x, \sin 2x\}$ is an orthonormal subset of the vector space $C[0, 2\pi]$ with respect to the inner product $\langle \mathbf{f}, \mathbf{g} \rangle = \int_0^{2\pi} f(x)g(x)\,dx$.

e. $\{1, \cos x, \sin x, \cos 2x, \sin 2x\}$ is a linearly independent subset of $C[0, 2\pi]$.

Chapter 6 Supplementary Exercises

1. Let R^4 have the Euclidean inner product.
   a. Find a vector in R^4 that is orthogonal to $\mathbf{u}_1 = (1, 0, 0, 0)$ and $\mathbf{u}_4 = (0, 0, 0, 1)$ and makes equal angles with the vectors $\mathbf{u}_2 = (0, 1, 0, 0)$ and $\mathbf{u}_3 = (0, 0, 1, 0)$.
   b. Find a vector $\mathbf{x} = (x_1, x_2, x_3, x_4)$ of length 1 that is orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_4$ above and such that the cosine of the angle between $\mathbf{x}$ and $\mathbf{u}_2$ is twice the cosine of the angle between $\mathbf{x}$ and $\mathbf{u}_3$.

2. Prove: If $\langle \mathbf{u}, \mathbf{v} \rangle$ is the Euclidean inner product on R^n, and if A is an $n \times n$ matrix, then
$$\langle \mathbf{u}, A\mathbf{v} \rangle = \langle A^T\mathbf{u}, \mathbf{v} \rangle$$
[*Hint:* Use the fact that $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = \mathbf{v}^T\mathbf{u}$.]

3. Let M_{22} have the inner product $\langle U, V \rangle = \text{tr}(U^TV) = \text{tr}(V^TU)$ that was defined in Example 6 of Section 6.1. Describe the orthogonal complement of
   a. the subspace of all diagonal matrices.
   b. the subspace of symmetric matrices.

4. Let $A\mathbf{x} = \mathbf{0}$ be a system of m equations in n unknowns. Show that
$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$
is a solution of this system if and only if $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is orthogonal to every row vector of A with respect to the Euclidean inner product on R^n.

5. Use the Cauchy–Schwarz inequality to show that if $a_1, a_2, \ldots, a_n$ are positive real numbers, then
$$(a_1 + a_2 + \cdots + a_n)\left(\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}\right) \geq n^2$$

6. Show that if $\mathbf{x}$ and $\mathbf{y}$ are vectors in an inner product space and c is any scalar, then
$$\|c\mathbf{x} + \mathbf{y}\|^2 = c^2\|\mathbf{x}\|^2 + 2c\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2$$

7. Let R^3 have the Euclidean inner product. Find two vectors of length 1 each of which is orthogonal to all three of the vectors $\mathbf{u}_1 = (1, 1, -1)$, $\mathbf{u}_2 = (-2, -1, 2)$, and $\mathbf{u}_3 = (-1, 0, 1)$.

8. Find a weighted Euclidean inner product on R^n such that the vectors

$$\begin{aligned}\mathbf{v}_1 &= (1, 0, 0, \dots, 0)\\ \mathbf{v}_2 &= (0, \sqrt{2}, 0, \dots, 0)\\ \mathbf{v}_3 &= (0, 0, \sqrt{3}, \dots, 0)\\ &\vdots\\ \mathbf{v}_n &= (0, 0, 0, \dots, \sqrt{n})\end{aligned}$$

form an orthonormal set.

9. Is there a weighted Euclidean inner product on R^2 for which the vectors $(1, 2)$ and $(3, -1)$ form an orthonormal set? Justify your answer.

10. If $\mathbf{u}$ and $\mathbf{v}$ are vectors in an inner product space V, then $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} - \mathbf{v}$ can be regarded as sides of a "triangle" in V (see the accompanying figure). Prove that the law of cosines holds for any such triangle; that is,

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

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

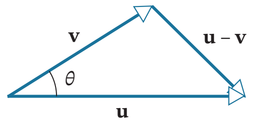


FIGURE Ex-10

11. a. As shown in Figure 3.2.6, the vectors $(k, 0, 0)$, $(0, k, 0)$, and $(0, 0, k)$ form the edges of a cube in R^3 with diagonal (k, k, k). Similarly, the vectors

$$(k, 0, 0, \dots, 0), \quad (0, k, 0, \dots, 0), \dots, \quad (0, 0, 0, \dots, k)$$

can be regarded as edges of a "cube" in R^n with diagonal $(k, k, k, \dots, k)$. Show that each of the above edges makes an angle of θ with the diagonal, where $\cos\theta = 1/\sqrt{n}$.

b. ***(Calculus required)*** What happens to the angle θ in part (a) as the dimension of R^n approaches ∞?

12. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space.

a. Prove that $\|\mathbf{u}\| = \|\mathbf{v}\|$ if and only if $\mathbf{u} + \mathbf{v}$ and $\mathbf{u} - \mathbf{v}$ are orthogonal.

b. Give a geometric interpretation of this result in R^2 with the Euclidean inner product.

13. Let $\mathbf{u}$ be a vector in an inner product space V, and let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be an orthonormal basis for V. Show that if α_i is the angle between $\mathbf{u}$ and $\mathbf{v}_i$, then

$$\cos^2\alpha_1 + \cos^2\alpha_2 + \cdots + \cos^2\alpha_n = 1$$

14. Prove: If $\langle \mathbf{u}, \mathbf{v}\rangle_1$ and $\langle \mathbf{u}, \mathbf{v}\rangle_2$ are two inner products on a vector space V, then the quantity $\langle \mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle_1 + \langle \mathbf{u}, \mathbf{v}\rangle_2$ is also an inner product.

15. Prove Theorem 6.2.5.

16. Prove: If A has linearly independent column vectors, and if $\mathbf{b}$ is orthogonal to the column space of A, then the least squares solution of $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = \mathbf{0}$.

17. Is there any value of s for which $x_1 = 1$ and $x_2 = 2$ is the least squares solution of the following linear system?

$$\begin{aligned} x_1 - x_2 &= 1\\ 2x_1 + 3x_2 &= 1\\ 4x_1 + 5x_2 &= s \end{aligned}$$

Explain your reasoning.

18. Show that if p and q are distinct positive integers, then the functions $f(x) = \sin px$ and $g(x) = \sin qx$ are orthogonal with respect to the inner product

$$\langle \mathbf{f}, \mathbf{g}\rangle = \int_0^{2\pi} f(x)g(x)\,dx$$

19. Show that if p and q are positive integers, then the functions $f(x) = \cos px$ and $g(x) = \sin qx$ are orthogonal with respect to the inner product

$$\langle \mathbf{f}, \mathbf{g}\rangle = \int_0^{2\pi} f(x)g(x)\,dx$$

20. Let W be the intersection of the planes

$$x + y + z = 0 \quad \text{and} \quad x - y + z = 0$$

in R^3. Find an equation for $W^{\perp}$.

21. Prove that if $ad - bc \neq 0$, then the matrix

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

has a unique QR-decomposition $A = QR$, where

$$Q = \frac{1}{\sqrt{a^2 + c^2}}\begin{bmatrix} a & -c\\ c & a\end{bmatrix}$$

$$R = \frac{1}{\sqrt{a^2 + c^2}}\begin{bmatrix} a^2 + c^2 & ab + cd\\ 0 & ad - bc\end{bmatrix}$$

CHAPTER 7

Diagonalization and Quadratic Forms

CHAPTER CONTENTS

Introduction

In Section 5.2 we found conditions that guaranteed the diagonalizability of an $n \times n$ matrix, but we did not consider which class or classes of matrices might actually satisfy those conditions. In this chapter we will show that every symmetric matrix is diagonalizable. This is an extremely important result because many applications utilize it in some essential way.

7.1 Orthogonal Matrices

In this section we will discuss the class of matrices whose inverses can be obtained by transposition. Such matrices occur in a variety of applications and arise as well as transition matrices when one orthonormal basis is changed to another.

Orthogonal Matrices

We begin with the following definition.

Definition 1

A square matrix A is said to be ***orthogonal*** if its transpose is the same as its inverse, that is, if

$$A^{-1} = A^T$$

or, equivalently, if

$$AA^T = A^TA = I \tag{1}$$

A matrix transformation $T_A: R^n \to R^n$ is said to be an ***orthogonal transformation*** or an ***orthogonal operator*** if A is an orthogonal matrix.

Recall from Theorem 1.6.3 and the related discussion that if either product in (1) holds, then so does the other. Thus, A is orthogonal if *either* $AA^T = I$ or $A^TA = I$.

EXAMPLE 1 | A 3×3 Orthogonal Matrix

The matrix

$$A = \begin{bmatrix} \frac{3}{7} & \frac{2}{7} & \frac{6}{7} \\ -\frac{6}{7} & \frac{3}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{6}{7} & -\frac{3}{7} \end{bmatrix}$$

is orthogonal since

$$A^TA = \begin{bmatrix} \frac{3}{7} & -\frac{6}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{3}{7} & \frac{6}{7} \\ \frac{6}{7} & \frac{2}{7} & -\frac{3}{7} \end{bmatrix} \begin{bmatrix} \frac{3}{7} & \frac{2}{7} & \frac{6}{7} \\ -\frac{6}{7} & \frac{3}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{6}{7} & -\frac{3}{7} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

EXAMPLE 2 | Rotation and Reflection Matrices Are Orthogonal

Recall from Table 5 of Section 1.8 that the standard matrix for the counterclockwise rotation about the origin of R^2 through an angle θ is

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

This matrix is orthogonal for all choices of θ since

$$A^TA = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

We leave it for you to verify that the reflection matrices in Tables 1 and 2 of Section 1.8 are all orthogonal.

Observe that for the orthogonal matrices in Examples 1 and 2, both the row vectors and the column vectors form orthonormal sets with respect to the Euclidean inner product. This is a consequence of the following theorem.

Theorem 7.1.1

The following are equivalent for an $n \times n$ matrix A.

(a) A is orthogonal.

(b) The row vectors of A form an orthonormal set in R^n with the Euclidean inner product.

(c) The column vectors of A form an orthonormal set in R^n with the Euclidean inner product.

Proof We will prove the equivalence of (a) and (b) and leave the equivalence of (a) and (c) as an exercise.

(a) ⇔ (b) Let $\mathbf{r}_i$ be the ith row vector and $\mathbf{c}_j$ the jth column vector of A. Since transposing a matrix converts its columns to rows and rows to columns, it follows that $\mathbf{c}_j^T = \mathbf{r}_j$. Thus, it follows from the row-column rule [Formula (5) of Section 1.3] and the bottom form listed in Table 1 of Section 3.2 that

$$AA^T = \begin{bmatrix} \mathbf{r}_1\mathbf{c}_1^T & \mathbf{r}_1\mathbf{c}_2^T & \cdots & \mathbf{r}_1\mathbf{c}_n^T \\ \mathbf{r}_2\mathbf{c}_1^T & \mathbf{r}_2\mathbf{c}_2^T & \cdots & \mathbf{r}_2\mathbf{c}_n^T \\ \vdots & \vdots & & \vdots \\ \mathbf{r}_n\mathbf{c}_1^T & \mathbf{r}_n\mathbf{c}_2^T & \cdots & \mathbf{r}_n\mathbf{c}_n^T \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1\cdot\mathbf{r}_1 & \mathbf{r}_1\cdot\mathbf{r}_2 & \cdots & \mathbf{r}_1\cdot\mathbf{r}_n \\ \mathbf{r}_2\cdot\mathbf{r}_1 & \mathbf{r}_2\cdot\mathbf{r}_2 & \cdots & \mathbf{r}_2\cdot\mathbf{r}_n \\ \vdots & \vdots & & \vdots \\ \mathbf{r}_n\cdot\mathbf{r}_1 & \mathbf{r}_n\cdot\mathbf{r}_2 & \cdots & \mathbf{r}_n\cdot r_n \end{bmatrix}$$

It is evident from this formula that $AA^T = I$ if and only if

$$\mathbf{r}_1\cdot\mathbf{r}_1 = \mathbf{r}_2\cdot\mathbf{r}_2 = \cdots = \mathbf{r}_n\cdot\mathbf{r}_n = 1$$

and

$$\mathbf{r}_i\cdot\mathbf{r}_j = 0 \quad \text{when } i \neq j$$

which are true if and only if $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n\}$ is an orthonormal set in R^n. ■

Warning Note that an orthogonal matrix has *orthonormal* rows and columns—not simply orthogonal rows and columns.

The following theorem lists four more fundamental properties of orthogonal matrices. The proofs are all straightforward and are left as exercises.

Theorem 7.1.2

(*a*) The transpose of an orthogonal matrix is orthogonal.
(*b*) The inverse of an orthogonal matrix is orthogonal.
(*c*) A product of orthogonal matrices is orthogonal.
(*d*) If A is orthogonal, then $\det(A) = 1$ or $\det(A) = -1$.

EXAMPLE 3 | $\det(A) = \pm 1$ for an Orthogonal Matrix A

The matrix

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

is orthogonal since its row (and column) vectors form orthonormal sets in R^2 with the Euclidean inner product. We leave it for you to verify that $\det(A) = 1$ and that interchanging the rows produces an orthogonal matrix whose determinant is -1.

Properties of Orthogonal Transformations

We observed in Example 2 that the standard matrices for the basic reflection and rotation operators on R^2 and R^3 are orthogonal. The next theorem will explain why this is so.

Theorem 7.1.3

If A is an $n \times n$ matrix, then the following are equivalent.

(*a*) A is orthogonal.
(*b*) $\|A\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ in R^n.
(*c*) $A\mathbf{x}\cdot A\mathbf{y} = \mathbf{x}\cdot\mathbf{y}$ for all $\mathbf{x}$ and $\mathbf{y}$ in R^n.

Proof We will prove the sequence of implications $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a)$.

(a) ⇒ (b) Assume that A is orthogonal, so that $A^TA = I$. It follows from Formula (26) of Section 3.2 that

$$\|A\mathbf{x}\| = (A\mathbf{x} \cdot A\mathbf{x})^{1/2} = (\mathbf{x} \cdot A^TA\mathbf{x})^{1/2} = (\mathbf{x} \cdot \mathbf{x})^{1/2} = \|\mathbf{x}\|$$

(b) ⇒ (c) Assume that $\|A\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ in R^n. From Theorem 3.2.7 we have

$$\begin{aligned} A\mathbf{x} \cdot A\mathbf{y} &= \tfrac{1}{4}\|A\mathbf{x} + A\mathbf{y}\|^2 - \tfrac{1}{4}\|A\mathbf{x} - A\mathbf{y}\|^2 = \tfrac{1}{4}\|A(\mathbf{x} + \mathbf{y})\|^2 - \tfrac{1}{4}\|A(\mathbf{x} - \mathbf{y})\|^2 \\ &= \tfrac{1}{4}\|\mathbf{x} + \mathbf{y}\|^2 - \tfrac{1}{4}\|\mathbf{x} - \mathbf{y}\|^2 = \mathbf{x} \cdot \mathbf{y} \end{aligned}$$

(c) ⇒ (a) Assume that $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}$ and $\mathbf{y}$ in R^n. It follows from Formula (26) of Section 3.2 that

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot A^TA\mathbf{y}$$

which can be rewritten as $\mathbf{x} \cdot (A^TA\mathbf{y} - \mathbf{y}) = 0$ or as

$$\mathbf{x} \cdot (A^TA - I)\mathbf{y} = 0$$

Since this equation holds for all $\mathbf{x}$ in R^n, it holds in particular if $\mathbf{x} = (A^TA - I)\mathbf{y}$, so

$$(A^TA - I)\mathbf{y} \cdot (A^TA - I)\mathbf{y} = 0$$

It follows from the positivity axiom for inner products that

$$(A^TA - I)\mathbf{y} = \mathbf{0}$$

Since this equation is satisfied by every vector $\mathbf{y}$ in R^n, it must be that $A^TA - I$ is the zero matrix (why?) and hence that $A^TA = I$. Thus, A is orthogonal. ■

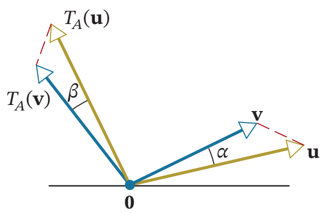


FIGURE 7.1.1

It follows from parts (a) and (b) of Theorem 7.1.3 that the orthogonal operators on R^n are precisely those operators that leave dot products and norms of vectors unchanged. However, as illustrated in **Figure 7.1.1**, this implies that orthogonal operators also leave angles and distances between vectors in R^n unchanged since these can be expressed in terms of norms [see Definition 2 and Formula (20) of Section 3.2].

Change of Orthonormal Basis

Orthonormal bases for inner product spaces are convenient because, as the following theorem shows, many familiar formulas hold for such bases. We leave the proof as an exercise.

Theorem 7.1.4

If S is an orthonormal basis for an n-dimensional inner product space V, and if

$$(\mathbf{u})_S = (u_1, u_2, \ldots, u_n) \quad \text{and} \quad (\mathbf{v})_S = (v_1, v_2, \ldots, v_n)$$

then:

(a) $\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$

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

(c) $\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 + u_2v_2 + \cdots + u_nv_n$

Remark Note that the three parts of Theorem 7.1.4 can be expressed as

$$\|\mathbf{u}\| = \|(\mathbf{u})_S\| \qquad d(\mathbf{u},\mathbf{v}) = d\big((\mathbf{u})_S,(\mathbf{v})_S\big) \qquad \langle \mathbf{u},\mathbf{v}\rangle = \big\langle (\mathbf{u})_S,(\mathbf{v})_S\big\rangle$$

where the norm, distance, and inner product on the left sides are relative to the inner product on V and on the right sides are relative to the Euclidean inner product on R^n. In short, norms, distances, and inner products of vectors in V can be computed from their coordinate vectors relative to an orthonormal basis using the Euclidean inner product.

Transitions between orthonormal bases for an inner product space are of special importance in geometry and various applications. The following theorem, whose proof is deferred to the end of this section, is concerned with transitions of this type.

Theorem 7.1.5

Let V be a finite-dimensional inner product space. If P is the transition matrix from one orthonormal basis for V to another orthonormal basis for V, then P is an orthogonal matrix.

EXAMPLE 4 | Rotation of Axes in 2-Space

In many problems a rectangular xy-coordinate system is given, and a new $x'y'$-coordinate system is obtained by rotating the xy-system counterclockwise about the origin through an angle θ. When this is done, each point Q in the plane has two sets of coordinates—coordinates (x,y) relative to the xy-system and coordinates (x',y') relative to the $x'y'$-system (**Figure 7.1.2*a***).

By introducing unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ along the positive x- and y-axes and unit vectors $\mathbf{u}_1'$ and $\mathbf{u}_2'$ along the positive x'- and y'-axes, we can regard this rotation as a change from an old basis $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ to a new basis $B' = \{\mathbf{u}_1', \mathbf{u}_2'\}$. Thus, with an appropriate adjustment in notation it follows from Formulas (7) and (8) of Section 4.7 that the new coordinates (x', y') and the old coordinates (x, y) of a point Q are related by the equation

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = P\begin{bmatrix} x \\ y \end{bmatrix} \tag{2}$$

where

$$P = \big[[\mathbf{u}_1]_{B'} \,|\, [\mathbf{u}_2]_{B'}\big]$$

Thus, to find P we must find the coordinates of the old basis vectors with respect to the new basis. We leave it for you to deduce the following results **Figure 7.1.2*b***.

$$[\mathbf{u}_1]_{B'} = \begin{bmatrix} \cos\theta \\ -\sin\theta \end{bmatrix} \quad \text{and} \quad [\mathbf{u}_2]_{B'} = \begin{bmatrix} \sin\theta \\ \cos\theta \end{bmatrix} \tag{3}$$

Thus

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \tag{4}$$

or equivalently

$$\boxed{\begin{aligned} x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta \end{aligned}} \tag{5}$$

These are sometimes called the ***rotation equations*** for R^2.

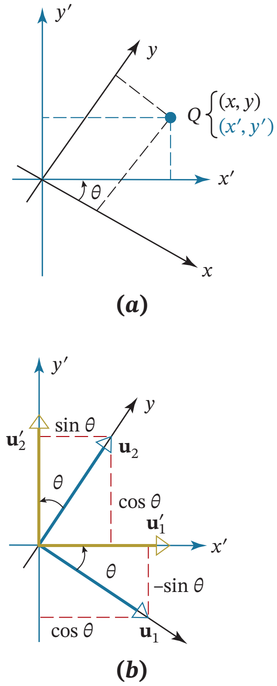


FIGURE 7.1.2

EXAMPLE 5 | Rotation of Axes in 2-Space

Use form (4) of the rotation equations for R^2 to find the new coordinates of the point $Q(2,1)$ if the coordinate axes of a rectangular coordinate system are rotated through an angle of $\theta = \pi/4$.

Solution Since

$$\sin\frac{\pi}{4} = \cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}$$

the equation in (4) becomes

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Thus, if the old coordinates of a point Q are $(x,y) = (2,-1)$, then

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{3}{\sqrt{2}} \end{bmatrix}$$

so the new coordinates of Q are $(x', y') = \left(\frac{1}{\sqrt{2}}, -\frac{3}{\sqrt{2}}\right)$.

Remark Observe that the coefficient matrix in (4) is the same as the standard matrix for the linear operator that rotates the vectors of R^2 through the angle $-\theta$ (see margin note for Table 5 of Section 1.8). This is to be expected since rotating the coordinate axes through the angle θ with the vectors of R^2 kept fixed has the same effect as rotating the vectors in R^2 through the angle $-\theta$ with the axes kept fixed.

EXAMPLE 6 | Rotation of Axes in 3-Space

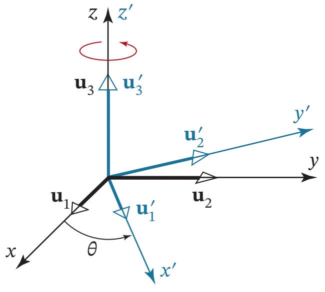


FIGURE 7.1.3

Suppose that a rectangular xyz-coordinate system is rotated around its z-axis counterclockwise (looking down the positive z-axis) through an angle θ (**Figure 7.1.3**). If we introduce unit vectors $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ along the positive x-, y-, and z-axes and unit vectors $\mathbf{u}_1'$, $\mathbf{u}_2'$, and $\mathbf{u}_3'$ along the positive x'-, y'-, and z'-axes, we can regard the rotation as a change from the old basis $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ to the new basis $B' = \{\mathbf{u}_1', \mathbf{u}_2', \mathbf{u}_3'\}$. In light of Example 4, it should be evident that

$$[\mathbf{u}_1']_B = \begin{bmatrix} \cos\theta \\ \sin\theta \\ 0 \end{bmatrix} \quad \text{and} \quad [\mathbf{u}_2']_B = \begin{bmatrix} -\sin\theta \\ \cos\theta \\ 0 \end{bmatrix}$$

Moreover, since $\mathbf{u}_3'$ extends 1 unit up the positive z'-axis,

$$[\mathbf{u}_3']_B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

It follows that the transition matrix from B' to B is

$$P = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and the transition matrix from B to B' is

$$P^{-1} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(verify). Thus, the new coordinates (x', y', z') of a point Q can be computed from its old coordinates (x, y, z) by

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

OPTIONAL: We conclude this section with an optional proof of Theorem 7.1.5.

Proof of Theorem 7.1.5 Assume that V is an n-dimensional inner product space and that P is the transition matrix from an orthonormal basis B' to an orthonormal basis B. We will denote the norm relative to the inner product on V by the symbol $\|\ \|_V$ to distinguish it from the norm relative to the Euclidean inner product on R^n, which we will denote by $\|\ \|$.

To prove that P is orthogonal, we will use Theorem 7.1.3 and show that $\|P\mathbf{x}\| = \|\mathbf{x}\|$ for every vector $\mathbf{x}$ in R^n. As a first step in this direction, recall from Theorem 7.1.4(a) that for any orthonormal basis for V the norm of any vector $\mathbf{u}$ in V is the same as the norm of its coordinate vector with respect to the Euclidean inner product, that is,

$$\|\mathbf{u}\|_V = \|[\mathbf{u}]_{B'}\| = \|[\mathbf{u}]_B\|$$

or

$$\|\mathbf{u}\|_V = \|[\mathbf{u}]_{B'}\| = \|P[\mathbf{u}]_{B'}\| \tag{6}$$

Now let $\mathbf{x}$ be any vector in R^n, and let $\mathbf{u}$ be the vector in V whose coordinate vector with respect to the basis B' is $\mathbf{x}$, that is, $[\mathbf{u}]_{B'} = \mathbf{x}$. Thus, from (6),

$$\|\mathbf{u}\| = \|\mathbf{x}\| = \|P\mathbf{x}\|$$

which proves that P is orthogonal. ■

Exercise Set 7.1

In each part of Exercises 1–4, determine whether the matrix is orthogonal, and if so find it inverse.

1. a. $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ **b.** $\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

2. a. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ **b.** $\begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$

3. a. $\begin{bmatrix} 0 & 1 & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$ **b.** $\begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$

4. a. $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{5}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & -\frac{5}{6} \\ \frac{1}{2} & \frac{1}{6} & -\frac{5}{6} & \frac{1}{6} \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 & 1 \\ 0 & \frac{1}{\sqrt{3}} & \frac{1}{2} & 0 \end{bmatrix}$

In Exercises 5–6, show that the matrix is orthogonal three ways: first by calculating A^TA, then by using part (b) of Theorem 7.1.1, and then by using part (c) of Theorem 7.1.1.

5. $A = \begin{bmatrix} \frac{4}{5} & 0 & -\frac{3}{5} \\ -\frac{9}{25} & \frac{4}{5} & -\frac{12}{25} \\ \frac{12}{25} & \frac{3}{5} & \frac{16}{25} \end{bmatrix}$ **6.** $A = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix}$

7. Let $T_A: R^3 \to R^3$ be multiplication by the orthogonal matrix in Exercise 5. Find $T_A(\mathbf{x})$ for the vector $\mathbf{x} = (-2, 3, 5)$, and confirm that $\|T_A(\mathbf{x})\| = \|\mathbf{x}\|$ relative to the Euclidean inner product on R^3.

8. Let $T_A: R^3 \to R^3$ be multiplication by the orthogonal matrix in Exercise 6. Find $T_A(\mathbf{x})$ for the vector $\mathbf{x} = (0, 1, 4)$, and confirm $\|T_A(\mathbf{x})\| = \|\mathbf{x}\|$ relative to the Euclidean inner product on R^3.

9. Are the standard matrices for the reflections in Tables 1 and 2 of Section 1.8 orthogonal?

10. Are the standard matrices for the orthogonal projections in Tables 3 and 4 of Section 1.8 orthogonal?

11. What conditions must a and b satisfy for the matrix

$$\begin{bmatrix} a+b & b-a \\ a-b & b+a \end{bmatrix}$$

to be orthogonal?

12. Under what conditions will a diagonal matrix be orthogonal?

13. Consider the rectangular $x'y'$-coordinate system obtained by rotating a rectangular xy-coordinate system counterclockwise through the angle $\theta = \pi/3$.

a. Find the $x'y'$-coordinates of the point whose xy-coordinates are $(-2, 6)$.

b. Find the xy-coordinates of the point whose $x'y'$-coordinates are $(5, 2)$.

14. Repeat Exercise 13 with $\theta = 3\pi/4$.

15. Consider the rectangular $x'y'z'$-coordinate system obtained by rotating a rectangular xyz-coordinate system counterclockwise about the z-axis (looking down the z-axis) through the angle $\theta = \pi/4$.

a. Find the $x'y'z'$-coordinates of the point whose xyz-coordinates are $(-1, 2, 5)$.

b. Find the xyz-coordinates of the point whose $x'y'z'$-coordinates are $(1, 6, -3)$.

16. Repeat Exercise 15 for a rotation of $\theta = 3\pi/4$ counterclockwise about the x-axis (looking along the positive x-axis toward the origin).

17. Repeat Exercise 15 for a rotation of $\theta = \pi/3$ counterclockwise about the y-axis (looking along the positive y-axis toward the origin).

18. A rectangular $x'y'z'$-coordinate system is obtained by rotating an xyz-coordinate system counterclockwise about the y-axis through an angle θ (looking along the positive y-axis toward the origin). Find a matrix A such that

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = A \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

where (x, y, z) and (x', y', z') are the coordinates of the same point in the xyz- and $x'y'z'$-systems, respectively.

19. Repeat Exercise 18 for a rotation about the x-axis.

20. A rectangular $x''y''z''$-coordinate system is obtained by first rotating a rectangular xyz-coordinate system 60° counterclockwise about the z-axis (looking down the positive z-axis) to obtain an $x'y'z'$-coordinate system, and then rotating the $x'y'z'$-coordinate system 45° counterclockwise about the y'-axis (looking along the positive y'-axis toward the origin). Find a matrix A such that

$$\begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = A \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

where (x, y, z) and (x'', y'', z'') are the xyz- and $x''y''z''$-coordinates of the same point.

21. A linear operator on R^2 is called ***rigid*** if it does not change the lengths of vectors, and it is called ***angle preserving*** if it does not change the angle between nonzero vectors.

a. Identify two different types of linear operators that are rigid.

b. Identify two different types of linear operators that are angle preserving.

c. Are there any linear operators on R^2 that are rigid and not angle preserving? Angle preserving and not rigid? Justify your answer.

22. Can an orthogonal operator $T_A: R^n \to R^n$ map nonzero vectors that are not orthogonal into orthogonal vectors? Justify your answer.

23. The set $S = \left\{\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{2}}x, \sqrt{\frac{3}{2}}x^2 - \sqrt{\frac{2}{3}}\right\}$ is an orthonormal basis for P_2 with respect to the evaluation inner product at the points $x_0 = -1$, $x_1 = 0$, $x_2 = 1$. Let $\mathbf{p} = p(x) = 1 + x + x^2$ and $\mathbf{q} = q(x) = 2x - x^2$.

a. Find $(\mathbf{p})_S$ and $(\mathbf{q})_S$.

b. Use Theorem 7.1.4 to compute $\|\mathbf{p}\|$, $d(\mathbf{p}, \mathbf{q})$ and $\langle \mathbf{p}, \mathbf{q} \rangle$.

24. The sets $S = \{1, x\}$ and $S' = \left\{\frac{1}{\sqrt{2}}(1+x), \frac{1}{\sqrt{2}}(1-x)\right\}$ are orthonormal bases for P_1 with respect to the standard inner product. Find the transition matrix P from S to S', and verify that the conclusion of Theorem 7.1.5 holds for P.

Working with Proofs

25. Prove that if $\mathbf{x}$ is an $n \times 1$ matrix, then the matrix

$$A = I_n - \frac{2}{\mathbf{x}^T\mathbf{x}}\mathbf{x}\mathbf{x}^T$$

is both orthogonal and symmetric.

26. Prove that a 2×2 orthogonal matrix A has only one of two possible forms:

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{or} \quad A = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

where $0 \le \theta < 2\pi$. [*Hint:* Start with a general 2×2 matrix A, and use the fact that the column vectors form an orthonormal set in R^2.]

27. **a.** Use the result in Exercise 26 to prove that multiplication by a 2×2 orthogonal matrix is a rotation if $\det(A) = 1$ and a reflection followed by a rotation if $\det(A) = -1$.

b. In the case where the transformation in part (a) is a reflection followed by a rotation, show that the same transformation can be accomplished by a single reflection about an appropriate line through the origin. What is that line? [*Hint:* See Formula (6) of Section 1.8.]

28. In each part, use the result in Exercise 27(a) to determine whether multiplication by A is a rotation or a reflection followed by rotation. Find the angle of rotation in both cases, and in the case where it is a reflection followed by a rotation find an equation for the line through the origin referenced in Exercise 27(b).

a. $A = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$ **b.** $A = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$

29. The result in Exercise 27(a) has an analog for 3×3 orthogonal matrices. It can be proved that multiplication by a 3×3 orthogonal matrix A is a rotation about some line through the origin of R^3 if $\det(A) = 1$ and is a reflection about some coordinate plane followed by a rotation about some line through the origin if $\det(A) = -1$. Use the first of these facts and Theorem 7.1.2 to prove that any composition of rotations about lines through the origin in R^3 can be accomplished by a single rotation about an appropriate line through the origin.

30. Prove the equivalence of statements (a) and (c) that are given in Theorem 7.1.1.

True-False Exercises

TF. In parts **(a)–(h)** determine whether the statement is true or false, and justify your answer.

a. The matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ is orthogonal.

b. The matrix $\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$ is orthogonal.

c. An $m \times n$ matrix A is orthogonal if $A^TA = I$.

d. A square matrix whose columns form an orthogonal set is orthogonal.

e. Every orthogonal matrix is invertible.

f. If A is an orthogonal matrix, then A^2 is orthogonal and $(\det A)^2 = 1$.

g. Every eigenvalue of an orthogonal matrix has absolute value 1.

h. If A is a square matrix and $\|A\mathbf{u}\| = 1$ for all unit vectors $\mathbf{u}$, then A is orthogonal.

Working with Technology

T1. If $\mathbf{a}$ is a nonzero vector in R^n, then $\mathbf{a}\mathbf{a}^T$ is called the ***outer product*** of $\mathbf{a}$ with itself, the subspace $\mathbf{a}^\perp$ is called the ***hyperplane*** in R^n orthogonal to $\mathbf{a}$, and the $n \times n$ orthogonal matrix

$$H_{\mathbf{a}^\perp} = I - \frac{2}{\mathbf{a}^T\mathbf{a}}\mathbf{a}\mathbf{a}^T$$

is called the ***Householder matrix*** or the ***Householder reflection*** about $\mathbf{a}^\perp$, named in honor of the American mathematician Alston S. Householder (1904–1993). In R^2 the matrix $H_{\mathbf{a}^\perp}$ represents a reflection about the line through the origin that is orthogonal to $\mathbf{a}$, and in R^3 it represents a reflection about the plane through the origin that is orthogonal to $\mathbf{a}$. In higher dimensions we can view $H_{\mathbf{a}^\perp}$ as a "reflection" about the hyperplane $\mathbf{a}^\perp$. Householder reflections are important in large-scale implementations of numerical algorithms, because they can be used to transform a given vector into a vector with specified zero components while leaving the other components unchanged. This is a consequence of the following theorem [see *Contemporary Linear Algebra*, by Howard Anton and Robert C. Busby (Hoboken, NJ: John Wiley & Sons, 2003, p. 422)].

Theorem

If $\mathbf{v}$ and $\mathbf{w}$ are distinct vectors in R^n with the same norm, then the Householder reflection about the hyperplane $(\mathbf{v} - \mathbf{w})^\perp$ maps $\mathbf{v}$ into $\mathbf{w}$ and conversely.

a. Find a Householder reflection that maps $\mathbf{v} = (4, 2, 4)$ into a vector $\mathbf{w}$ that has zeros as its second and third components. Find $\mathbf{w}$.

b. Find a Householder reflection that maps $\mathbf{v} = (3, 4, 2, 4)$ into the vector whose last two entries are zero, while leaving the first entry unchanged. Find $\mathbf{w}$.

7.2 Orthogonal Diagonalization

In this section we will be concerned with the problem of diagonalizing a symmetric matrix A. As we will see, this problem is closely related to that of finding an orthonormal basis for R^n that consists of eigenvectors of A. Problems of this type are important because many of the matrices that arise in applications are symmetric.

The Orthogonal Diagonalization Problem

In Section 5.2 we defined two square matrices, A and B, to be *similar* if there is an *invertible* matrix P such that $P^{-1}AP = B$. In this section we will be concerned with the special case in which it is possible to find an *orthogonal* matrix P for which this relationship holds.

We begin with the following definition.

Definition 1

If A and B are square matrices, then we say that B is ***orthogonally similar*** to A if there is an orthogonal matrix P such that $B = P^TAP$.

Note that if B is orthogonally similar to A, then it is also true that A is orthogonally similar to B since we can express A as $A = PBP^T = Q^TBQ$, where $Q = P^T$. This being the case we will say that A and B are ***orthogonally similar matrices*** if either is orthogonally similar to the other.

If A is orthogonally similar to some diagonal matrix, say

$$P^TAP = D$$

then we say A is ***orthogonally diagonalizable*** and P ***orthogonally diagonalizes*** A.

Our first goal in this section is to determine what conditions a matrix must satisfy to be orthogonally diagonalizable. As an initial step, observe that there is no hope of orthogonally diagonalizing a matrix that is not symmetric. To see why this is so, suppose that

$$P^TAP = D \tag{1}$$

where P is an orthogonal matrix and D is a diagonal matrix. Multiplying the left side of (1) by P, the right side by P^T, and then using the fact that $PP^T = P^TP = I$, we can rewrite this equation as

$$A = PDP^T \tag{2}$$

Now transposing both sides of this equation and using the fact that a diagonal matrix is the same as its transpose we obtain

$$A^T = (PDP^T)^T = (P^T)^TD^TP^T = PDP^T = A$$

so A must be symmetric if it is orthogonally diagonalizable.

Conditions for Orthogonal Diagonalizability

We showed above that in order for a square matrix A to be orthogonally diagonalizable it must be symmetric. Our next theorem will show that the converse is true if A has *real entries* and the orthogonality is with respect to the Euclidean inner product on R^n.

Theorem 7.2.1

If A is an $n \times n$ matrix with real entries, then the following are equivalent.

(*a*) A is orthogonally diagonalizable.

(*b*) A has an orthonormal set of n eigenvectors.

(*c*) A is symmetric.

***Proof* (*a*) ⇒ (*b*)** Since A is orthogonally diagonalizable, there is an orthogonal matrix P such that $P^{-1}AP$ is diagonal. As shown in Formula (2) in the proof of Theorem 5.2.1, the n column vectors of P are eigenvectors of A. Since P is orthogonal, these column vectors are orthonormal, so A has n orthonormal eigenvectors.

(*b*) ⇒ (*a*) Assume that A has an orthonormal set of n eigenvectors $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$. As shown in the proof of Theorem 5.2.1, the matrix P with these eigenvectors as columns diagonalizes A. Since these eigenvectors are orthonormal, the matrix P is orthogonal and thus orthogonally diagonalizes A.

(*a*) ⇒ (*c*) In the proof that (*a*) ⇒ (*b*) we showed that an orthogonally diagonalizable $n \times n$ matrix A is orthogonally diagonalized by an $n \times n$ matrix P whose columns form an orthonormal set of eigenvectors of A. Let D be the diagonal matrix

$$D = P^TAP$$

from which it follows that

$$A = PDP^T$$

Thus,

$$A^T = (PDP^T)^T = PD^TP^T = PDP^T = A$$

which shows that A is symmetric.

(*c*) ⇒ (*a*) The proof of this part is beyond the scope of this text. However, because it is such an important result we have outlined the structure of its proof in the exercises. ■

Properties of Symmetric Matrices

Our next goal is to devise a procedure for orthogonally diagonalizing a symmetric matrix, but before we can do so, we need the following critical theorem about eigenvalues and eigenvectors of symmetric matrices.

Theorem 7.2.2

If A is a symmetric matrix with real entries, then:

(*a*) The eigenvalues of A are all real numbers.

(*b*) Eigenvectors from different eigenspaces are orthogonal.

Part (*a*), which requires results about complex vector spaces, will be discussed in Section 7.5.

Proof (b) Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors corresponding to distinct eigenvalues λ_1 and λ_2 of the matrix A. We want to show that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. Our proof of this involves the trick of starting with the expression $A\mathbf{v}_1 \cdot \mathbf{v}_2$. It follows from Formula (26) of Section 3.2 and the symmetry of A that

$$A\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1 \cdot A^T\mathbf{v}_2 = \mathbf{v}_1 \cdot A\mathbf{v}_2 \tag{3}$$

But $\mathbf{v}_1$ is an eigenvector of A corresponding to λ_1, and $\mathbf{v}_2$ is an eigenvector of A corresponding to λ_2, so (3) yields the relationship

$$\lambda_1\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1 \cdot \lambda_2\mathbf{v}_2$$

which can be rewritten as

$$(\lambda_1 - \lambda_2)(\mathbf{v}_1 \cdot \mathbf{v}_2) = 0 \tag{4}$$

But $\lambda_1 - \lambda_2 \neq 0$, since λ_1 and λ_2 were assumed distinct, so it follows from (4) that

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = 0 \quad \blacksquare$$

Theorem 7.2.2 yields the following procedure for orthogonally diagonalizing a symmetric matrix.

Orthogonally Diagonalizing an $n \times n$ Symmetric Matrix

Step 1. Find a basis for each eigenspace of A.

Step 2. Apply the Gram–Schmidt process to each of these bases to obtain an orthonormal basis for each eigenspace.

Step 3. Form the matrix P whose columns are the vectors constructed in Step 2. This matrix will orthogonally diagonalize A, and the eigenvalues on the diagonal of $D = P^TAP$ will be in the same order as their corresponding eigenvectors in P.

Remark The justification of this procedure should be clear: Theorem 7.2.2 ensures that eigenvectors from *different* eigenspaces are orthogonal, and applying the Gram–Schmidt process ensures that the eigenvectors within the *same* eigenspace are orthonormal. Thus the *entire* set of eigenvectors obtained by this procedure will be orthonormal.

EXAMPLE 1 | Orthogonally Diagonalizing a Symmetric Matrix

Find an orthogonal matrix P that diagonalizes

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

Solution We leave it for you to verify that the characteristic equation of A is

$$\det(\lambda I - A) = \det\begin{bmatrix} \lambda - 4 & -2 & -2 \\ -2 & \lambda - 4 & -2 \\ -2 & -2 & \lambda - 4 \end{bmatrix} = (\lambda - 2)^2(\lambda - 8) = 0$$

Thus, the distinct eigenvalues of A are $\lambda = 2$ and $\lambda = 8$. By the method used in Example 7 of Section 5.1, it can be shown that

$$\mathbf{u}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \tag{5}$$

form a basis for the eigenspace corresponding to $\lambda = 2$. Applying the Gram–Schmidt process to $\{\mathbf{u}_1, \mathbf{u}_2\}$ yields the following orthonormal eigenvectors (verify):

$$\mathbf{v}_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix} \tag{6}$$

The eigenspace corresponding to $\lambda = 8$ has

$$\mathbf{u}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

as a basis. Applying the Gram–Schmidt process to $\{\mathbf{u}_3\}$ (i.e., normalizing $\mathbf{u}_3$) yields

$$\mathbf{v}_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

Finally, using $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ as column vectors, we obtain

$$P = \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

which orthogonally diagonalizes A. As a check, we leave it for you to confirm that

$$P^TAP = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{bmatrix}$$

Spectral Decomposition

If A is a symmetric matrix with real entries that is orthogonally diagonalized by

$$P = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n]$$

and if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A corresponding to the unit eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$, then we know that $D = P^TAP$, where D is a diagonal matrix with the eigenvalues in the diagonal positions. It follows from this that the matrix A can be expressed as

$$A = PDP^T = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix}$$

$$= [\lambda_1\mathbf{u}_1 \quad \lambda_2\mathbf{u}_2 \quad \cdots \quad \lambda_n\mathbf{u}_n] \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix}$$

Multiplying out, we obtain the formula

$$\boxed{A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T} \tag{7}$$

which is called a ***spectral decomposition of A***.*

Note that each term of the spectral decomposition of A has the form $\lambda \mathbf{u}\mathbf{u}^T$, where $\mathbf{u}$ is a unit eigenvector of A in column form, and λ is an eigenvalue of A corresponding to $\mathbf{u}$. Since $\mathbf{u}$ has size $n \times 1$, it follows that the product $\mathbf{u}\mathbf{u}^T$ has size $n \times n$. It can be proved (though we will not do it) that $\mathbf{u}\mathbf{u}^T$ is the standard matrix for the orthogonal projection of R^n on the subspace spanned by the vector $\mathbf{u}$. Accepting this to be so, the spectral decomposition of A states that the image of a vector $\mathbf{x}$ under multiplication by a symmetric matrix A can be obtained by projecting $\mathbf{x}$ orthogonally on the lines (one-dimensional subspaces) determined by the eigenvectors of A, then scaling those projections by the eigenvalues, and then adding the scaled projections. Here is an example.

EXAMPLE 2 | A Geometric Interpretation of a Spectral Decomposition

The matrix

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

has eigenvalues $\lambda_1 = -3$ and $\lambda_2 = 2$ with corresponding eigenvectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

(verify). Normalizing these basis vectors yields

$$\mathbf{u}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \frac{\mathbf{x}_2}{\|\mathbf{x}_2\|} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$$

so a spectral decomposition of A is

$$\begin{aligned}\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix} = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T &= (-3)\begin{bmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix} + (2)\begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}\begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \\ &= (-3)\begin{bmatrix} \frac{1}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{4}{5} \end{bmatrix} + (2)\begin{bmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{bmatrix}\end{aligned} \tag{8}$$

where, as noted above, the 2×2 matrices on the right side of (8) are the standard matrices for the orthogonal projections onto the eigenspaces corresponding to the eigenvalues $\lambda_1 = -3$ and $\lambda_2 = 2$, respectively.

Now let us see what this spectral decomposition tells us about the image of the vector $\mathbf{x} = (1, 1)$ under multiplication by A. Writing $\mathbf{x}$ in column form, it follows that

$$A\mathbf{x} = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \tag{9}$$

*The terminology *spectral decomposition* is derived from the fact that the set of all eigenvalues of a matrix A is sometimes called the ***spectrum*** of A. The terminology *eigenvalue decomposition* is due to Professor Dan Kalman, who introduced it in an award-winning paper entitled "A Singularly Valuable Decomposition: The SVD of a Matrix," *The College Mathematics Journal*, Vol. 27, No. 1, January 1996.

and from (8) that

$$A\mathbf{x} = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = (-3)\begin{bmatrix} \frac{1}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{4}{5} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} + (2)\begin{bmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= (-3)\begin{bmatrix} -\frac{1}{5} \\ \frac{2}{5} \end{bmatrix} + (2)\begin{bmatrix} \frac{6}{5} \\ \frac{3}{5} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{3}{5} \\ -\frac{6}{5} \end{bmatrix} + \begin{bmatrix} \frac{12}{5} \\ \frac{6}{5} \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \tag{10}$$

Formulas (9) and (10) provide two different ways of viewing the image of the vector $(1, 1)$ under multiplication by A: Formula (9) tells us directly that the image of this vector is $(3, 0)$, whereas Formula (10) tells us that this image can also be obtained by projecting $(1, 1)$ onto the eigenspaces corresponding to $\lambda_1 = -3$ and $\lambda_2 = 2$ to obtain the vectors $\left(-\frac{1}{5}, \frac{2}{5}\right)$ and $\left(\frac{6}{5}, \frac{3}{5}\right)$, then scaling by the eigenvalues to obtain $\left(\frac{3}{5}, -\frac{6}{5}\right)$ and $\left(\frac{12}{5}, \frac{6}{5}\right)$, and then adding these vectors (see **Figure 7.2.1**).

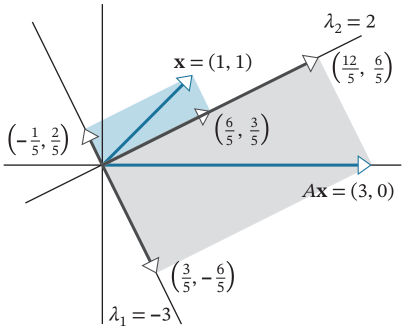


FIGURE 7.2.1

The Nondiagonalizable Case

If A is an $n \times n$ matrix that is not orthogonally diagonalizable, it may still be possible to achieve considerable simplification in the form of P^TAP by choosing the orthogonal matrix P appropriately. We will consider two theorems (without proof) that illustrate this. The first, due to the German mathematician Issai Schur, states that every square matrix A is orthogonally similar to an upper triangular matrix that has the eigenvalues of A on the main diagonal.

Theorem 7.2.3

Schur's Theorem

If A is an $n \times n$ matrix with real entries and real eigenvalues, then there is an orthogonal matrix P such that P^TAP is an upper triangular matrix of the form

$$P^TAP = \begin{bmatrix} \lambda_1 & \times & \times & \cdots & \times \\ 0 & \lambda_2 & \times & \cdots & \times \\ 0 & 0 & \lambda_3 & \cdots & \times \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{11}$$

in which $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A repeated according to multiplicity.

Historical Note

Issai Schur (1875–1941)

The life of the German mathematician Issai Schur is a sad reminder of the effect that Nazi policies had on Jewish intellectuals during the 1930s. Schur was a brilliant mathematician and a popular lecturer who attracted many students and researchers to the University of Berlin, where he worked and taught. His lectures sometimes attracted so many students that opera glasses were needed to see him from the back row. Schur's life became increasingly difficult under Nazi rule, and in April of 1933 he was forced to "retire" from the university under a law that prohibited non-Aryans from holding "civil service" positions. There was an outcry from many of his students and colleagues who respected and liked him, but it did not stave off his complete dismissal in 1935. Schur, who thought of himself as a loyal German, never understood the persecution and humiliation he received at Nazi hands. He left Germany for Palestine in 1939, a broken man. Lacking in financial resources, he had to sell his beloved mathematics books and lived in poverty until his death in 1941.

[Image: Courtesy Electronic Publishing Services, Inc., New York City]

It is common to denote the upper triangular matrix in (11) by S (for Schur), in which case that equation would be rewritten as

$$A = PSP^T \tag{12}$$

which is called a ***Schur decomposition*** of A.

The next theorem, due to the German mathematician and electrical engineer Karl Hessenberg (1904–1959), states that every square matrix with real entries is orthogonally similar to a matrix in which each entry below the first ***subdiagonal*** is zero (Figure 7.2.2). Such a matrix is said to be in ***upper Hessenberg form***.

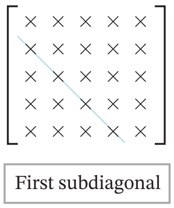


FIGURE 7.2.2

Theorem 7.2.4

Hessenberg's Theorem

If A is an $n \times n$ matrix with real entries, then there is an orthogonal matrix P such that P^TAP is a matrix of the form

$$P^TAP = \begin{bmatrix} \times & \times & \cdots & \times & \times & \times \\ \times & \times & \cdots & \times & \times & \times \\ 0 & \times & \ddots & \times & \times & \times \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \times & \times & \times \\ 0 & 0 & \cdots & 0 & \times & \times \end{bmatrix} \tag{13}$$

Note that unlike those in (11), the diagonal entries in (13) are usually *not* the eigenvalues of A.

It is common to denote the upper Hessenberg matrix in (13) by H (for Hessenberg), in which case that equation can be rewritten as

$$A = PHP^T \tag{14}$$

which is called an ***upper Hessenberg decomposition*** of A.

Remark In many numerical algorithms the initial matrix is first converted to upper Hessenberg form to reduce the amount of computation in subsequent parts of the algorithm. Many computer packages have built-in commands for finding Schur and Hessenberg decompositions.

Exercise Set 7.2

In Exercises 1–6, find the characteristic equation of the given symmetric matrix, and then by inspection determine the dimensions of the eigenspaces.

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

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

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

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

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

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

In Exercises 7–14, find a matrix P that orthogonally diagonalizes A, and determine $P^{-1}AP$.

7. $A = \begin{bmatrix} 6 & 2\sqrt{3} \\ 2\sqrt{3} & 7 \end{bmatrix}$

8. $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$

9. $A = \begin{bmatrix} -2 & 0 & -36 \\ 0 & -3 & 0 \\ -36 & 0 & -23 \end{bmatrix}$

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

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

12. $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

13. $A = \begin{bmatrix} -7 & 24 & 0 & 0 \\ 24 & 7 & 0 & 0 \\ 0 & 0 & -7 & 24 \\ 0 & 0 & 24 & 7 \end{bmatrix}$

14. $A = \begin{bmatrix} 3 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

In Exercises 15–18, find the spectral decomposition of the matrix.

15. $\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$

16. $\begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix}$

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

18. $\begin{bmatrix} -2 & 0 & -36 \\ 0 & -3 & 0 \\ -36 & 0 & -23 \end{bmatrix}$

In Exercises 19–20, determine whether there exists a 3×3 symmetric matrix whose eigenvalues are $\lambda_1 = -1, \lambda_2 = 3, \lambda_3 = 7$ and for which the corresponding eigenvectors are as stated. If there is such a matrix, find it, and if there is none, explain why not.

19. $\mathbf{x}_1 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

20. $\mathbf{x}_1 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

21. Let A be a diagonalizable matrix with the property that eigenvectors corresponding to distinct eigenvalues are orthogonal. Must A be symmetric? Explain your reasoning.

22. Assuming that $b \neq 0$, find a matrix that orthogonally diagonalizes

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix}$$

23. Let $T_A\colon R^2 \to R^2$ be multiplication by A. Find two orthogonal unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ such that $T_A(\mathbf{u}_1)$ and $T_A(\mathbf{u}_2)$ are orthogonal.

a. $A = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

24. Let $T_A\colon R^3 \to R^3$ be multiplication by A. Find two orthogonal unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ such that $T_A(\mathbf{u}_1)$ and $T_A(\mathbf{u}_2)$ are orthogonal.

a. $A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$

Working with Proofs

25. Prove that if A is any $m \times n$ matrix, then A^TA has an orthonormal set of n eigenvectors.

26. Prove: If $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is an orthonormal basis for R^n, and if A can be expressed as

$$A = c_1\mathbf{u}_1\mathbf{u}_1^T + c_2\mathbf{u}_2\mathbf{u}_2^T + \cdots + c_n\mathbf{u}_n\mathbf{u}_n^T$$

then A is symmetric and has eigenvalues $c_1, c_2, \ldots, c_n$.

27. Use the result in Exercise 29 of Section 5.1 to prove Theorem 7.2.2 (a) for 2×2 symmetric matrices.

28. a. Prove that if $\mathbf{v}$ is any $n \times 1$ matrix and I is the $n \times n$ identity matrix, then $I - \mathbf{v}\mathbf{v}^T$ is orthogonally diagonalizable.

b. Find a matrix P that orthogonally diagonalizes $I - \mathbf{v}\mathbf{v}^T$ if

$$\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

29. Prove that if A is a symmetric orthogonal matrix, then 1 and -1 are the only possible eigenvalues.

30. Is the converse of Theorem 7.2.2 (b) true? Justify your answer.

31. In this exercise we will show that a symmetric matrix A is orthogonally diagonalizable, thereby completing the missing part of Theorem 7.2.1. We will proceed in two steps: first we will show that A is diagonalizable, and then we will build on that result to show that A is orthogonally diagonalizable.

a. Assume that A is a symmetric $n \times n$ matrix. One way to prove that A is diagonalizable is to show that for each eigenvalue λ_0 the geometric multiplicity is equal to the algebraic multiplicity. For this purpose, assume that the geometric multiplicity of λ_0 is k, let $B_0 = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be an orthonormal basis for the eigenspace corresponding to the eigenvalue λ_0, extend this to an orthonormal basis $B_0 = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ for R^n, and let P be the matrix having the vectors of B as columns. As shown in Exercise 41(b) of Section 5.2, the product AP can be written as

$$AP = P\begin{bmatrix} \lambda_0 I_k & X \\ 0 & Y \end{bmatrix}$$

Use the fact that B is an orthonormal basis to prove that $X = 0$ [a zero matrix of size $n \times (n-k)$].

b. It follows from part (a) and Exercise 41(c) of Section 5.2 that A has the same characteristic polynomial as

$$C = P\begin{bmatrix} \lambda_0 I_k & 0 \\ 0 & Y \end{bmatrix}$$

Use this fact and Exercise 41(d) of Section 5.2 to prove that the algebraic multiplicity of λ_0 is the same as the geometric multiplicity of λ_0. This establishes that A is diagonalizable.

c. Use Theorem 7.2.2(b) and the fact that A is diagonalizable to prove that A is orthogonally diagonalizable.

True-False Exercises

TF. In parts **(a)–(g)** determine whether the statement is true or false, and justify your answer.

a. If A is a square matrix, then AA^T and A^TA are orthogonally diagonalizable.

b. If $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors from distinct eigenspaces of a symmetric matrix with real entries, then

$$\|\mathbf{v}_1 + \mathbf{v}_2\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2$$

c. Every orthogonal matrix is orthogonally diagonalizable.

d. If A is both invertible and orthogonally diagonalizable, then A^{-1} is orthogonally diagonalizable.

e. Every eigenvalue of an orthogonal matrix has absolute value 1.

f. If A is an $n \times n$ orthogonally diagonalizable matrix, then there exists an orthonormal basis for R^n consisting of eigenvectors of A.

g. If A is orthogonally diagonalizable, then A has real eigenvalues.

Working with Technology

T1. If your technology utility has an orthogonal diagonalization capability, use it to confirm the final result obtained in Example 1.

T2. For the given matrix A, find orthonormal bases for the eigenspaces of A, and use those basis vectors to construct an orthogonal matrix P for which P^TAP is diagonal.

$$A = \begin{bmatrix} -4 & 2 & -2 \\ 2 & -7 & 4 \\ -2 & 4 & -7 \end{bmatrix}$$

T3. Find a spectral decomposition of the matrix A in Exercise T2.

7.3 Quadratic Forms

In this section we will use matrix methods to study real-valued functions of several variables in which each term is either the square of a variable or the product of two variables. Such functions arise in a variety of applications, including geometry, vibrations of mechanical systems, statistics, and electrical engineering.

Definition of a Quadratic Form

Expressions of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

occurred in our study of linear equations and linear systems. If $a_1, a_2, \ldots, a_n$ are treated as constants, then this expression is a real-valued function of the ***variables*** $x_1, x_2, \ldots, x_n$ and is called a ***linear form*** on R^n. All variables in a linear form occur to the first power and there are no products of variables. Here we will be concerned with ***quadratic forms*** on R^n, which are functions of the form

$$a_1x_1^2 + a_2x_2^2 + \cdots + a_nx_n^2 + (\text{all possible terms } a_kx_ix_j \text{ in which } i \neq j)$$

The terms of the form $a_kx_ix_j$ in which i is $\neq j$ are called ***cross product terms***. It is common to combine the cross product terms involving x_ix_j with those involving x_jx_i to avoid duplication. Thus, a general quadratic form on R^2 would typically be expressed as

$$a_1x_1^2 + a_2x_2^2 + 2a_3x_1x_2 \tag{1}$$

and a general quadratic form on R^3 as

$$a_1x_1^2 + a_2x_2^2 + a_3x_3^2 + 2a_4x_1x_2 + 2a_5x_1x_3 + 2a_6x_2x_3 \tag{2}$$

If, as usual, we do not distinguish between the number a and the 1×1 matrix $[a]$, and if we let $\mathbf{x}$ be the column vector of variables, then (1) and (2) can be expressed in matrix form as

$$[x_1 \quad x_2]\begin{bmatrix} a_1 & a_3 \\ a_3 & a_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{x}^T A\mathbf{x}$$

$$[x_1 \quad x_2 \quad x_3]\begin{bmatrix} a_1 & a_4 & a_5 \\ a_4 & a_2 & a_6 \\ a_5 & a_6 & a_3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{x}^T A\mathbf{x}$$

(verify). Note that the matrix A in these formulas is symmetric, that its diagonal entries are the coefficients of the squared terms, and its off-diagonal entries are half the coefficients of the cross product terms. In general, if A is a symmetric $n \times n$ matrix and $\mathbf{x}$ is an $n \times 1$ column vector of variables, then we call the function

$$Q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} \tag{3}$$

the ***quadratic form associated with A***. When convenient, (3) can be expressed in dot product notation as

$$\boxed{\mathbf{x}^T A\mathbf{x} = \mathbf{x} \cdot A\mathbf{x} = A\mathbf{x} \cdot \mathbf{x}} \tag{4}$$

In the case where A is a diagonal matrix, the quadratic form $\mathbf{x}^T A\mathbf{x}$ has no cross product terms; for example, if A has diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$, then

$$\mathbf{x}^T A\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2$$

EXAMPLE 1 | Expressing Quadratic Forms in Matrix Notation

In each part, express the quadratic form in the matrix notation $\mathbf{x}^T A\mathbf{x}$, where A is symmetric.

(*a*) $2x^2 + 6xy - 5y^2$ (*b*) $x_1^2 + 7x_2^2 - 3x_3^2 + 4x_1x_2 - 2x_1x_3 + 8x_2x_2$

Solution The diagonal entries of A are the coefficients of the squared terms, and the off-diagonal entries are half the coefficients of the cross product terms, so

$$2x^2 + 6xy - 5y^2 = [x \quad y]\begin{bmatrix} 2 & 3 \\ 3 & -5 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

$$x_1^2 + 7x_2^2 - 3x_3^2 + 4x_1x_2 - 2x_1x_3 + 8x_2x_3 = [x_1 \quad x_2 \quad x_3]\begin{bmatrix} 1 & 2 & -1 \\ 2 & 7 & 4 \\ -1 & 4 & -3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Change of Variable in a Quadratic Form

There are three important kinds of problems that occur in applications of quadratic forms:

Problem 1 If $\mathbf{x}^T A\mathbf{x}$ is a quadratic form on R^2 or R^3, what kind of curve or surface is represented by the equation $\mathbf{x}^T A\mathbf{x} = k$?

Problem 2 If $\mathbf{x}^T A\mathbf{x}$ is a quadratic form on R^n, what conditions must A satisfy for $\mathbf{x}^T A\mathbf{x}$ to have positive values for $\mathbf{x} \neq \mathbf{0}$?

Problem 3 If $\mathbf{x}^T A\mathbf{x}$ is a quadratic form on R^n, what are its maximum and minimum values if $\mathbf{x}$ is constrained to satisfy $\|\mathbf{x}\| = 1$?

We will consider the first two problems in this section and the third problem in the next.

Many of the techniques for solving these problems are based on simplifying the quadratic form $\mathbf{x}^T A\mathbf{x}$ by making a substitution

$$\mathbf{x} = P\mathbf{y} \tag{5}$$

that expresses the variables $x_1, x_2, \ldots, x_n$ in terms of new variables $y_1, y_2, \ldots, y_n$. If P is invertible, then we call (5) a ***change of variable***, and if P is orthogonal, then we call (5) an ***orthogonal change of variable***.

The following result, called the ***Principal Axes Theorem***, shows that by making an appropriate orthogonal change of variable in a quadratic form it is possible to eliminate its cross product terms, thereby producing a simpler quadratic form that is generally easier to work with.

Theorem 7.3.1

The Principal Axes Theorem

If A is a symmetric $n \times n$ matrix, then there is an orthogonal change of variable that transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into a quadratic form $\mathbf{y}^T D\mathbf{y}$ with no cross product terms. Specifically, if P orthogonally diagonalizes A, then making the change of variable $\mathbf{x} = P\mathbf{y}$ in the quadratic form $\mathbf{x}^T A\mathbf{x}$ yields the quadratic form

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

in which $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A corresponding to the eigenvectors that form the successive columns of P.

Proof If we make the change of variable $\mathbf{x} = P\mathbf{y}$ in the quadratic form $\mathbf{x}^T A\mathbf{x}$, then we obtain

$$\mathbf{x}^T A\mathbf{x} = (P\mathbf{y})^T A(P\mathbf{y}) = \mathbf{y}^T P^T AP\mathbf{y} = \mathbf{y}^T(P^T AP)\mathbf{y} \tag{6}$$

Since the matrix $B = P^T AP$ is symmetric (verify), the effect of the change of variable is to produce a new quadratic form $\mathbf{y}^T B\mathbf{y}$ in the variables $y_1, y_2, \ldots, y_n$. In particular, if we choose P to orthogonally diagonalize A, then the new quadratic form will be $\mathbf{y}^T D\mathbf{y}$, where D is a diagonal matrix with the eigenvalues of A on the main diagonal; that is,

$$\begin{aligned}\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} &= [y_1 \quad y_2 \quad \cdots \quad y_n]\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \quad \blacksquare\end{aligned}$$

EXAMPLE 2 | An Illustration of the Principal Axes Theorem

Find an orthogonal change of variable that eliminates the cross product terms in the quadratic form $Q = x_1^2 - x_3^2 - 4x_1x_2 + 4x_2x_3$, and express Q in terms of the new variables.

Solution The quadratic form can be expressed in matrix notation as

$$Q = \mathbf{x}^T A\mathbf{x} = [x_1 \quad x_2 \quad x_3]\begin{bmatrix} 1 & -2 & 0 \\ -2 & 0 & 2 \\ 0 & 2 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

The characteristic equation of the matrix A is

$$\begin{vmatrix} \lambda - 1 & 2 & 0 \\ 2 & \lambda & -2 \\ 0 & -2 & \lambda + 1 \end{vmatrix} = \lambda^3 - 9\lambda = \lambda(\lambda + 3)(\lambda - 3) = 0$$

so the eigenvalues are $\lambda = 0, -3, 3$. We leave it for you to show that orthonormal bases for the three eigenspaces are

$$\lambda = 0: \quad \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{bmatrix}, \quad \lambda = -3: \quad \begin{bmatrix} -\frac{1}{3} \\ -\frac{2}{3} \\ \frac{2}{3} \end{bmatrix}, \quad \lambda = 3: \quad \begin{bmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$$

Thus, a substitution $\mathbf{x} = P\mathbf{y}$ that eliminates the cross product terms is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

This produces the new quadratic form

$$Q = \mathbf{y}^T(P^TAP)\mathbf{y} = [y_1 \quad y_2 \quad y_3] \begin{bmatrix} 0 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = -3y_2^2 + 3y_3^2$$

in which there are no cross product terms.

Remark If A is a symmetric $n \times n$ matrix, then the quadratic form $\mathbf{x}^TA\mathbf{x}$ is a real-valued function whose range is the set of all possible values for $\mathbf{x}^TA\mathbf{x}$ as $\mathbf{x}$ varies over R^n. It can be shown that an orthogonal change of variable $\mathbf{x} = P\mathbf{y}$ does not alter the range of a quadratic form; that is, the set of all values for $\mathbf{x}^TA\mathbf{x}$ as $\mathbf{x}$ varies over R^n is the same as the set of all values for $\mathbf{y}^T(P^TAP)\mathbf{y}$ as $\mathbf{y}$ varies over R^n.

Quadratic Forms in Geometry

Recall that a ***conic section*** or ***conic*** is a curve that results by cutting a double-napped cone with a plane (**Figure 7.3.1**). The most important conic sections are ellipses, parabolas, and hyperbolas, which result when the cutting plane does not pass through the vertex. Circles are special cases of ellipses that result when the cutting plane is perpendicular to the axis of symmetry of the cone. If the cutting plane passes through the vertex, then the resulting intersection is called a ***degenerate conic***. The possibilities are a point, a pair of intersecting lines, or a single line.

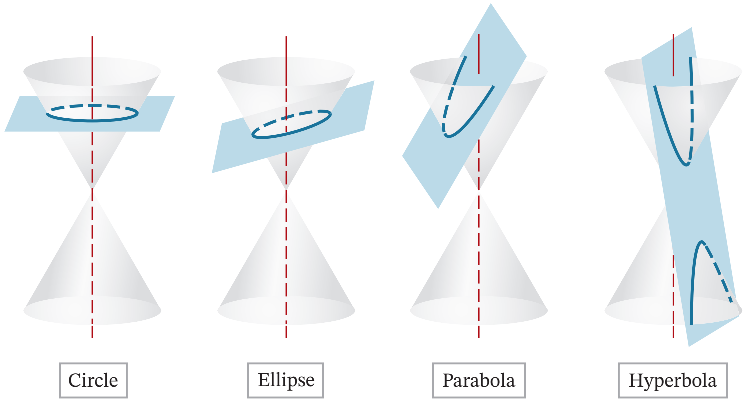


FIGURE 7.3.1

Quadratic forms in R^2 arise naturally in the study of conic sections. For example, it is shown in analytic geometry that an equation of the form

$$ax^2 + 2bxy + cy^2 + dx + ey + f = 0 \tag{7}$$

in which a, b, and c are not all zero, represents a conic section.* If $d = e = 0$ in (7), then there are no linear terms, so the equation becomes

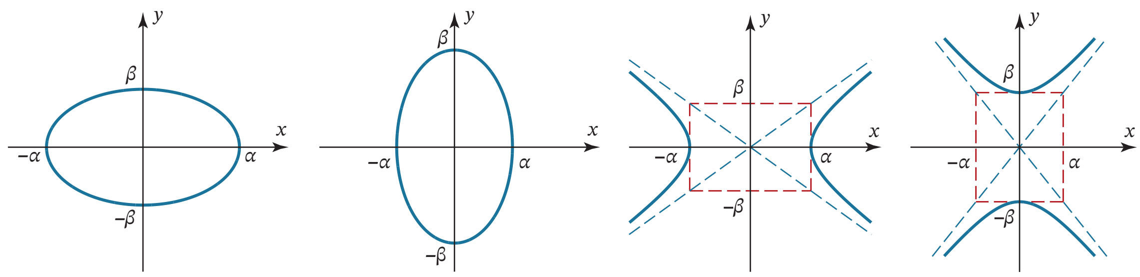
$$ax^2 + 2bxy + cy^2 + f = 0 \tag{8}$$

and is said to represent a ***central conic***. These include circles, ellipses, and hyperbolas, but not parabolas. Furthermore, if $b = 0$ in (8), then there is no cross product term (i.e., term involving xy), and the equation

$$ax^2 + cy^2 + f = 0 \tag{9}$$

is said to represent a ***central conic in standard position***. The most important conics of this type are shown in **Table 1**.

TABLE 1 Central Conics in Standard Position

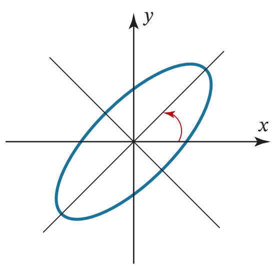


$\dfrac{x^2}{\alpha^2} + \dfrac{y^2}{\beta^2} = 1$ $(\alpha \ge \beta > 0)$	$\dfrac{x^2}{\alpha^2} + \dfrac{y^2}{\beta^2} = 1$ $(\beta \ge \alpha > 0)$	$\dfrac{x^2}{\alpha^2} - \dfrac{y^2}{\beta^2} = 1$ $(\alpha > 0, \beta > 0)$	$\dfrac{y^2}{\beta^2} - \dfrac{x^2}{\alpha^2} = 1$ $(\alpha > 0, \beta > 0)$

If we take the constant f in Equations (8) and (9) to the right side and let $k = -f$, then we can rewrite these equations in matrix form as

$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = k \quad \text{and} \quad \begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} a & 0 \\ 0 & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = k \tag{10}$$

The first of these corresponds to Equation (8) in which there is a cross product term $2bxy$, and the second corresponds to Equation (9) in which there is no cross product term. Geometrically, the existence of a cross product term signals that the graph of the quadratic form is rotated about the origin, as in **Figure 7.3.2**. The three-dimensional analogs of the equations in (10) are



A central conic rotated out of standard position

FIGURE 7.3.2

$$\begin{bmatrix} x & y & z \end{bmatrix}\begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = k \quad \text{and} \quad \begin{bmatrix} x & y & z \end{bmatrix}\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = k \tag{11}$$

If a, b, and c are not all zero, then the graphs in R^3 of the equations in (11) are called ***central quadrics***; the graph of the second of these equations, which is a special case of the first, is called a ***central quadric in standard position***.

*We must also allow for the possibility that there are no real values of x and y that satisfy the equation, as with $x^2 + y^2 + 1 = 0$. In such cases we say that the equation has ***no graph*** or has ***an empty graph***.

Identifying Conic Sections

We are now ready to consider the first of the three problems posed earlier, identifying the curve or surface represented by an equation $\mathbf{x}^T A\mathbf{x} = k$ in two or three variables. We will focus on the two-variable case. We noted earlier that an equation of the form

$$ax^2 + 2bxy + cy^2 + f = 0 \tag{12}$$

represents a central conic. If $b = 0$, then the conic is in standard position, and if $b \neq 0$, it is rotated. It is an easy matter to identify central conics in standard position by matching the equation with one of the standard forms. For example, the equation

$$9x^2 + 16y^2 - 144 = 0$$

can be rewritten as

$$\frac{x^2}{16} + \frac{y^2}{9} = 1$$

which, by comparison with Table 1, is the ellipse shown in **Figure 7.3.3**.

If a central conic is rotated out of standard position, then it can be identified by first rotating the coordinate axes to put it in standard position and then matching the resulting equation with one of the standard forms in Table 1. To find a rotation that eliminates the cross product term in the equation

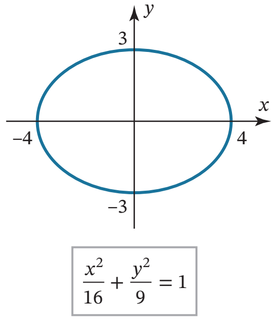


FIGURE 7.3.3

$$ax^2 + 2bxy + cy^2 = k \tag{13}$$

it will be convenient to express the equation in the matrix form

$$\mathbf{x}^T A\mathbf{x} = [x \quad y]\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = k \tag{14}$$

and look for a change of variable

$$\mathbf{x} = P\mathbf{x}'$$

that diagonalizes A and for which $\det(P) = 1$. Since we saw in Example 4 of Section 7.1 that the transition matrix

$$P = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{15}$$

has the effect of rotating the xy-axes of a rectangular coordinate system through an angle θ, our problem reduces to finding θ that diagonalizes A, thereby eliminating the cross product term in (13). If we make this change of variable, then in the $x'y'$-coordinate system, Equation (14) will become

$$\mathbf{x}'^T D\mathbf{x}' = [x' \quad y']\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} = k \tag{16}$$

where λ_1 and λ_2 are the eigenvalues of A. The conic can now be identified by writing (16) in the form

$$\lambda_1 x'^2 + \lambda_2 y'^2 = k \tag{17}$$

and performing the necessary algebra to match it with one of the standard forms in Table 1. For example, if λ_1, λ_2, and k are positive, then (17) represents an ellipse with an axis of length $2\sqrt{k/\lambda_1}$ in the x'-direction and $2\sqrt{k/\lambda_2}$ in the y'-direction. The first column vector of P, which is a unit eigenvector corresponding to λ_1, is along the positive x'-axis; and the second column vector of P, which is a unit eigenvector corresponding to λ_2, is a unit vector along the y'-axis. These are called the ***principal axes*** of the ellipse, which explains why Theorem 7.3.1 is called "the Principal Axes Theorem." (See **Figure 7.3.4**.)

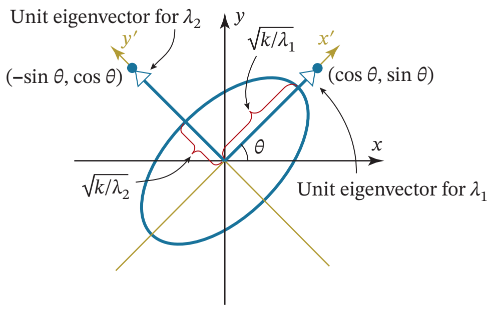


FIGURE 7.3.4

EXAMPLE 3 | Identifying a Conic by Eliminating the Cross Product Term

(*a*) Identify the conic whose equation is $5x^2 - 4xy + 8y^2 - 36 = 0$ by rotating the xy-axes to put the conic in standard position.

(*b*) Find the angle θ through which you rotated the xy-axes in part (a).

Solution (*a*) The given equation can be written in the matrix form

$$\mathbf{x}^T A\mathbf{x} = 36$$

where

$$A = \begin{bmatrix} 5 & -2 \\ -2 & 8 \end{bmatrix}$$

The characteristic polynomial of A is

$$\begin{vmatrix} \lambda - 5 & 2 \\ 2 & \lambda - 8 \end{vmatrix} = (\lambda - 4)(\lambda - 9)$$

so the eigenvalues are $\lambda = 4$ and $\lambda = 9$. We leave it for you to show that orthonormal bases for the eigenspaces are

$$\lambda = 4: \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}, \quad \lambda = 9: \begin{bmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$$

Thus, A is orthogonally diagonalized by

$$P = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \tag{18}$$

Had it turned out that

$$\det(P) = -1$$

then we would have interchanged the columns of P to reverse the sign.

Moreover, it happens by chance that $\det(P) = 1$, so we are assured that the substitution $\mathbf{x} = P\mathbf{x}'$ performs a rotation of axes. It follows from (16) that the equation of the conic in the $x'y'$-coordinate system is

$$\begin{bmatrix} x' & y' \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = 36$$

which we can write as

$$4x'^2 + 9y'^2 = 36 \quad \text{or} \quad \frac{x'^2}{9} + \frac{y'^2}{4} = 1$$

We can now see from Table 1 that the conic is an ellipse whose axis has length $2\alpha = 6$ in the x'-direction and length $2\beta = 4$ in the y'-direction.

Solution (*b*) It follows from (15) that

$$P = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

which implies that

$$\cos\theta = \frac{2}{\sqrt{5}}, \quad \sin\theta = \frac{1}{\sqrt{5}}, \quad \tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{1}{2}$$

Thus, $\theta = \tan^{-1}\frac{1}{2} \approx 26.6°$ (**Figure 7.3.5**).

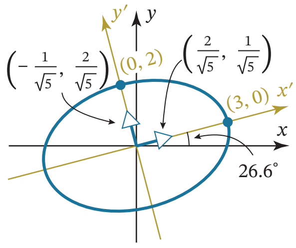


FIGURE 7.3.5

Remark In the exercises we will ask you to show that if $b \neq 0$, then the cross product term in the equation

$$ax^2 + 2bxy + cy^2 = k$$

can be eliminated by a rotation through an angle θ that satisfies

$$\boxed{\cot 2\theta = \frac{a-c}{2b}} \tag{19}$$

We leave it for you to confirm that this is consistent with part (b) of the last example.

Positive Definite Quadratic Forms

We will now consider the second of the two problems posed earlier, determining conditions under which $\mathbf{x}^T A\mathbf{x} > 0$ for all nonzero values of $\mathbf{x}$. We will explain why this is important shortly, but first let us introduce some terminology.

Definition 1

A quadratic form $\mathbf{x}^T A\mathbf{x}$ is said to be

positive definite if $\mathbf{x}^T A\mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$;

negative definite if $\mathbf{x}^T A\mathbf{x} < 0$ for $\mathbf{x} \neq \mathbf{0}$;

indefinite if $\mathbf{x}^T A\mathbf{x}$ has both positive and negative values.

The terminology in Definition 1 also applies to the matrix A; that is, A is positive definite, negative definite, or indefinite in accordance with whether the associated quadratic form has that property.

The following theorem, whose proof is deferred to the end of the section, provides a way of using eigenvalues to determine whether a matrix A and its associated quadratic form $\mathbf{x}^T A\mathbf{x}$ are positive definite, negative definite, or indefinite.

Theorem 7.3.2

If A is a symmetric matrix, then:

(*a*) $\mathbf{x}^T A\mathbf{x}$ is positive definite if and only if all eigenvalues of A are positive.

(*b*) $\mathbf{x}^T A\mathbf{x}$ is negative definite if and only if all eigenvalues of A are negative.

(*c*) $\mathbf{x}^T A\mathbf{x}$ is indefinite if and only if A has at least one positive eigenvalue and at least one negative eigenvalue.

Remark The three classifications in Definition 1 do not exhaust all possibilities. Specifically:

- $\mathbf{x}^T A\mathbf{x}$ is ***positive semidefinite*** if $\mathbf{x}^T A\mathbf{x} \geq 0$ if $\mathbf{x} \neq \mathbf{0}$
- $\mathbf{x}^T A\mathbf{x}$ is ***negative semidefinite*** if $\mathbf{x}^T A\mathbf{x} \leq 0$ if $\mathbf{x} \neq \mathbf{0}$

Observe that every positive definite form is positive semidefinite, but not conversely, and every negative definite form is negative semidefinite, but not conversely. By adjusting the proof of Theorem 7.3.2 (given at the end of this section) appropriately, one can show that if all eigenvalues of A are nonnegative, then $\mathbf{x}^T A\mathbf{x}$ is positive semidefinite, and if they are all nonpositive then $\mathbf{x}^T A\mathbf{x}$ is negative semidefinite.

EXAMPLE 4 | Positive Definite Quadratic Forms

It is not usually possible to tell from the signs of the entries in a symmetric matrix A whether that matrix is positive definite, negative definite, or indefinite. For example, the entries of the matrix

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

are nonnegative, but the matrix is indefinite since its eigenvalues are $\lambda = 1, 4, -2$ (verify). To see this another way, write out the quadratic form as

$$\mathbf{x}^T A\mathbf{x} = [x_1 \quad x_2 \quad x_3] \begin{bmatrix} 3 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 3x_1^2 + 2x_1x_2 + 2x_1x_3 + 4x_2x_3$$

We can now see, for example, that

$$\mathbf{x}^T A\mathbf{x} = 4 \quad \text{for} \quad x_1 = 0, \quad x_2 = 1, \quad x_3 = 1$$

and

$$\mathbf{x}^T A\mathbf{x} = -4 \quad \text{for} \quad x_1 = 0, \quad x_2 = 1, \quad x_3 = -1$$

Positive definite and negative definite matrices are invertible. Why?

Classifying Conic Sections Using Eigenvalues

If $\mathbf{x}^T B\mathbf{x} = k$ is the equation of a conic, and if $k \neq 0$, then we can divide through by k and rewrite the equation in the form

$$\mathbf{x}^T A\mathbf{x} = 1 \tag{20}$$

where $A = (1/k)B$. If we now rotate the coordinate axes to eliminate the cross product term (if any) in this equation, then the equation of the conic in the new coordinate system will be of the form

$$\lambda_1 x'^2 + \lambda_2 y'^2 = 1 \tag{21}$$

in which λ_1 and λ_2 are the eigenvalues of A. The particular type of conic represented by this equation will depend on the signs of the eigenvalues λ_1 and λ_2. For example, you should be able to see from (21) that:

- $\mathbf{x}^T A\mathbf{x} = 1$ represents an ellipse if $\lambda_1 > 0$ and $\lambda_2 > 0$.
- $\mathbf{x}^T A\mathbf{x} = 1$ has no graph if $\lambda_1 < 0$ and $\lambda_2 < 0$.
- $\mathbf{x}^T A\mathbf{x} = 1$ represents a hyperbola if λ_1 and λ_2 have opposite signs.

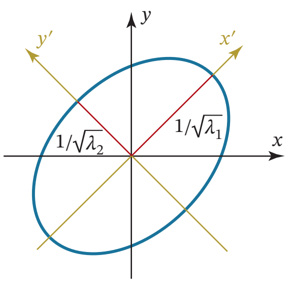


FIGURE 7.3.6

In the case of the ellipse, Equation (21) can be rewritten as

$$\frac{x'^2}{(1/\sqrt{\lambda_1})^2} + \frac{y'^2}{(1/\sqrt{\lambda_2})^2} = 1 \tag{22}$$

so the axes of the ellipse have lengths $2/\sqrt{\lambda_1}$ and $2/\sqrt{\lambda_2}$ (**Figure 7.3.6**).

The following theorem is an immediate consequence of this discussion and Theorem 7.3.2.

Theorem 7.3.3

If A is a symmetric 2×2 matrix, then:

(*a*) $\mathbf{x}^T A\mathbf{x} = 1$ represents an ellipse if A is positive definite.

(*b*) $\mathbf{x}^T A\mathbf{x} = 1$ has no graph if A is negative definite.

(*c*) $\mathbf{x}^T A\mathbf{x} = 1$ represents a hyperbola if A is indefinite.

In Example 3 we performed a rotation to show that the equation

$$5x^2 - 4xy + 8y^2 - 36 = 0$$

represents an ellipse with a major axis of length 6 and a minor axis of length 4. This conclusion can also be obtained by rewriting the equation in the form

$$\tfrac{5}{36}x^2 - \tfrac{1}{9}xy + \tfrac{2}{9}y^2 = 1$$

and showing that the associated matrix

$$A = \begin{bmatrix} \frac{5}{36} & -\frac{1}{18} \\ -\frac{1}{18} & \frac{2}{9} \end{bmatrix}$$

has eigenvalues $\lambda_1 = \frac{1}{9}$ and $\lambda_2 = \frac{1}{4}$. These eigenvalues are positive, so the matrix A is positive definite and the equation represents an ellipse. Moreover, it follows from (21) that the axes of the ellipse have lengths $2/\sqrt{\lambda_1} = 6$ and $2/\sqrt{\lambda_2} = 4$, which is consistent with Example 3.

Identifying Positive Definite Matrices

As positive definite matrices arise in many applications, it will be useful to learn a little more about them. We already know that a symmetric matrix is positive definite if and only if its eigenvalues are all positive; now we will give a criterion that can be used to determine whether a symmetric matrix is positive definite without the need for finding the eigenvalues. For this purpose we define the ***k*th *principal submatrix*** of an $n \times n$ matrix A to be the $k \times k$ submatrix consisting of the first k rows and columns of A. For example, here are the principal submatrices of a general 4×4 matrix:

$$\left[\begin{array}{c:ccc} a_{11} & a_{12} & a_{13} & a_{14} \\ \hdashline a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right] \quad \left[\begin{array}{cc:cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hdashline a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right] \quad \left[\begin{array}{ccc:c} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ \hdashline a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right] \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

First principal submatrix | Second principal submatrix | Third principal submatrix | Fourth principal submatrix = A

The following theorem, which we state without proof, provides a determinant test for ascertaining whether a symmetric matrix is positive definite.

Theorem 7.3.4

If A is a symmetric matrix, then:

(a) A is positive definite if and only if the determinant of every principal submatrix is positive.

(b) A is negative definite if and only if the determinants of the principal submatrices alternate between negative and positive values starting with a negative value for the determinant of the first principal submatrix.

(c) A is indefinite if and only if it is neither positive definite nor negative definite and at least one principal submatrix has a positive determinant and at least one has a negative determinant.

EXAMPLE 5 | Working with Principal Submatrices

The matrix

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

is positive definite since the determinants

$$|2| = 2, \quad \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3, \quad \begin{vmatrix} 2 & -1 & -3 \\ -1 & 2 & 4 \\ -3 & 4 & 9 \end{vmatrix} = 1$$

are all positive. Thus, we are guaranteed that all eigenvalues of A are positive and $\mathbf{x}^T A\mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$.

OPTIONAL: We conclude this section with an optional proof of Theorem 7.3.2.

Proofs of Theorem 7.3.2(a) and (b) It follows from the principal axes theorem (Theorem 7.3.1) that there is an orthogonal change of variable $\mathbf{x} = P\mathbf{y}$ for which

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \tag{23}$$

where the λ's are the eigenvalues of A. Moreover, it follows from the invertibility of P that $\mathbf{y} \neq \mathbf{0}$ if and only if $\mathbf{x} \neq \mathbf{0}$, so the values of $\mathbf{x}^T A\mathbf{x}$ for $\mathbf{x} \neq \mathbf{0}$ are the same as the values of $\mathbf{y}^T D\mathbf{y}$ for $\mathbf{y} \neq \mathbf{0}$. Thus, it follows from (23) that $\mathbf{x}^T A\mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$ if and only if all of the λ's in that equation are positive, and that $\mathbf{x}^T A\mathbf{x} < 0$ for $\mathbf{x} \neq \mathbf{0}$ if and only if all of the λ's are negative. This proves parts (a) and (b).

Proof (c) Assume that A has at least one positive eigenvalue and at least one negative eigenvalue, and to be specific, suppose that $\lambda_1 > 0$ and $\lambda_2 < 0$ in (23). Then

$$\mathbf{x}^T A\mathbf{x} > 0 \quad \text{if} \quad y_1 = 1 \text{ and all other } y\text{'s are } 0$$

and

$$\mathbf{x}^T A\mathbf{x} < 0 \quad \text{if} \quad y_2 = 1 \text{ and all other } y\text{'s are } 0$$

which proves that $\mathbf{x}^T A\mathbf{x}$ is indefinite. Conversely, if $\mathbf{x}^T A\mathbf{x} > 0$ for some $\mathbf{x}$, then $\mathbf{y}^T D\mathbf{y} > 0$ for some $\mathbf{y}$, so at least one of the λ's in (23) must be positive. Similarly, if $\mathbf{x}^T A\mathbf{x} < 0$ for some $\mathbf{x}$, then $\mathbf{y}^T D\mathbf{y} < 0$ for some $\mathbf{y}$, so at least one of the λ's in (23) must be negative, which completes the proof. ■

Exercise Set 7.3

In Exercises 1–2, express the quadratic form in the matrix notation $\mathbf{x}^T A\mathbf{x}$, *where* A *is a symmetric matrix.*

1. a. $3x_1^2 + 7x_2^2$ **b.** $4x_1^2 - 9x_2^2 - 6x_1x_2$

c. $9x_1^2 - x_2^2 + 4x_3^2 + 6x_1x_2 - 8x_1x_3 + x_2x_3$

2. a. $5x_1^2 + 5x_1x_2$ **b.** $-7x_1x_2$

c. $x_1^2 + x_2^2 - 3x_3^2 - 5x_1x_2 + 9x_1x_3$

In Exercises 3–4, find a formula for the quadratic form that does not use matrices.

3. $[x \quad y]\begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$

4. $[x_1 \quad x_2 \quad x_3]\begin{bmatrix} -2 & \frac{7}{2} & 1 \\ \frac{7}{2} & 0 & 6 \\ 1 & 6 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

In Exercises 5–8, find an orthogonal change of variables that eliminates the cross product terms in the quadratic form Q, *and express* Q *in terms of the new variables.*

5. $Q = 2x_1^2 + 2x_2^2 - 2x_1x_2$

6. $Q = 5x_1^2 + 2x_2^2 + 4x_3^2 + 4x_1x_2$

7. $Q = 3x_1^2 + 4x_2^2 + 5x_3^2 + 4x_1x_2 - 4x_2x_3$

8. $Q = 2x_1^2 + 5x_2^2 + 5x_3^2 + 4x_1x_2 - 4x_1x_3 - 8x_2x_3$

In Exercises 9–10, express the quadratic equation in the matrix form $\mathbf{x}^T A\mathbf{x} + K\mathbf{x} + f = 0$, *where* $\mathbf{x}^T A\mathbf{x}$ *is the associated quadratic form and* K *is an appropriate matrix.*

9. a. $2x^2 + xy + x - 6y + 2 = 0$

b. $y^2 + 7x - 8y - 5 = 0$

10. a. $x^2 - xy + 5x + 8y - 3 = 0$

b. $5xy = 8$

In Exercises 11–12, identify the conic section represented by the equation.

11. a. $2x^2 + 5y^2 = 20$ **b.** $x^2 - y^2 - 8 = 0$

c. $7y^2 - 2x = 0$ **d.** $x^2 + y^2 - 25 = 0$

12. a. $4x^2 + 9y^2 = 1$ **b.** $4x^2 - 5y^2 = 20$

c. $-x^2 = 2y$ **d.** $x^2 - 3 = -y^2$

In Exercises 13–16, identify the conic section represented by the equation by rotating axes to place the conic in standard position. Find an equation of the conic in the rotated coordinates, and find the angle of rotation.

13. $2x^2 - 4xy - y^2 + 8 = 0$ **14.** $5x^2 + 4xy + 5y^2 = 9$

15. $11x^2 + 24xy + 4y^2 - 15 = 0$ **16.** $x^2 + xy + y^2 = \frac{1}{2}$

In Exercises 17–18, determine by inspection whether the matrix is positive definite, negative definite, indefinite, positive semidefinite, or negative semidefinite.

17. a. $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ **b.** $\begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}$ **c.** $\begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}$

d. $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ **e.** $\begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix}$

18. a. $\begin{bmatrix} 2 & 0 \\ 0 & -5 \end{bmatrix}$ **b.** $\begin{bmatrix} -2 & 0 \\ 0 & -5 \end{bmatrix}$ **c.** $\begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}$

d. $\begin{bmatrix} 0 & 0 \\ 0 & -5 \end{bmatrix}$ **e.** $\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$

In Exercises 19–24, classify the quadratic form as positive definite, negative definite, indefinite, positive semidefinite, or negative semidefinite.

19. $x_1^2 + x_2^2$ **20.** $-x_1^2 - 3x_2^2$ **21.** $(x_1 - x_2)^2$

22. $-(x_1 - x_2)^2$ **23.** $x_1^2 - x_2^2$ **24.** x_1x_2

In Exercises 25–26, show that the matrix A *is positive definite first by using Theorem 7.3.2 and then by using Theorem 7.3.4.*

25. a. $A = \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

26. a. $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{bmatrix}$

In Exercises 27–28, use Theorem 7.3.4 to classify the matrix as positive definite, negative definite, or indefinite.

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

28. a. $A = \begin{bmatrix} 4 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$ **b.** $A = \begin{bmatrix} -4 & -1 & 1 \\ -1 & -2 & -1 \\ 1 & -1 & -2 \end{bmatrix}$

In Exercises 29–30, find all values of k for which the quadratic form is positive definite.

29. $5x_1^2 + x_2^2 + kx_3^2 + 4x_1x_2 - 2x_1x_3 - 2x_2x_3$

30. $3x_1^2 + x_2^2 + 2x_3^2 - 2x_1x_3 + 2kx_2x_3$

31. Let $\mathbf{x}^T A\mathbf{x}$ be a quadratic form in the variables $x_1, x_2, \ldots, x_n$, and define $T: R^n \to R$ by $T(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$.

a. Show that $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + 2\mathbf{x}^T A\mathbf{y} + T(\mathbf{y})$.

b. Show that $T(c\mathbf{x}) = c^2 T(\mathbf{x})$.

32. Express the quadratic form $(c_1x_1 + c_2x_2 + \cdots + c_nx_n)^2$ in the matrix notation $\mathbf{x}^T A\mathbf{x}$, where A is symmetric.

33. In statistics, the quantities

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

$$s_x^2 = \frac{1}{n-1}\left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\right]$$

(cont.)

are called, respectively, the ***sample mean*** and ***sample variance*** of $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

a. Express the quadratic form s_x^2 in the matrix notation $\mathbf{x}^T A\mathbf{x}$, where A is symmetric.

b. Is s_x^2 a positive definite quadratic form? Explain.

34. The graph in an xyz-coordinate system of an equation of form $ax^2 + by^2 + cz^2 = 1$ in which a, b, and c are positive is a surface called a ***central ellipsoid in standard position*** (see the accompanying figure). This is the three-dimensional generalization of the ellipse $ax^2 + by^2 = 1$ in the xy-plane. The intersections of the ellipsoid $ax^2 + by^2 + cz^2 = 1$ with the coordinate axes determine three line segments called the ***axes*** of the ellipsoid. If a central ellipsoid is rotated about the origin so two or more of its axes do not coincide with any of the coordinate axes, then the resulting equation will have one or more cross product terms.

a. Show that the equation

$$\tfrac{4}{3}x^2 + \tfrac{4}{3}y^2 + \tfrac{4}{3}z^2 + \tfrac{4}{3}xy + \tfrac{4}{3}xz + \tfrac{4}{3}yz = 1$$

represents an ellipsoid, and find the lengths of its axes. [*Suggestion:* Write the equation in the form $\mathbf{x}^T A\mathbf{x} = 1$ and make an orthogonal change of variable to eliminate the cross product terms.]

b. What property must a symmetric 3×3 matrix have in order for the equation $\mathbf{x}^T A\mathbf{x} = 1$ to represent an ellipsoid?

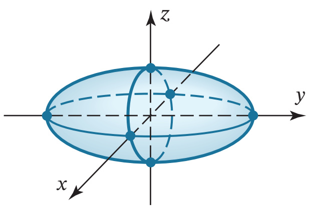


FIGURE Ex-34

35. What property must a symmetric 2×2 matrix A have for $\mathbf{x}^T A\mathbf{x} = 1$ to represent a circle?

Working with Proofs

36. Prove: If $b \neq 0$, then the cross product term can be eliminated from the quadratic form $ax^2 + 2bxy + cy^2$ by rotating the coordinate axes through an angle θ that satisfies the equation

$$\cot 2\theta = \frac{a - c}{2b}$$

37. Prove: If A is an $n \times n$ symmetric matrix all of whose eigenvalues are nonnegative, then $\mathbf{x}^T A\mathbf{x} \geq 0$ for all nonzero $\mathbf{x}$ in the vector space R^n.

True-False Exercises

TF. In parts (a)–(l) determine whether the statement is true or false, and justify your answer.

a. If all eigenvalues of a symmetric matrix A are positive, then A is positive definite.

b. $x_1^2 - x_2^2 + x_3^2 + 4x_1x_2x_3$ is a quadratic form.

c. $(x_1 - 3x_2)^2$ is a quadratic form.

d. A positive definite matrix is invertible.

e. A symmetric matrix is either positive definite, negative definite, or indefinite.

f. If A is positive definite, then $-A$ is negative definite.

g. $\mathbf{x} \cdot \mathbf{x}$ is a quadratic form for all $\mathbf{x}$ in R^n.

h. If A is symmetric and invertible, and if $\mathbf{x}^T A\mathbf{x}$ is a positive definite quadratic form, then $\mathbf{x}^T A^{-1}\mathbf{x}$ is also a positive definite quadratic form.

i. If A is symmetric and has only positive eigenvalues, then $\mathbf{x}^T A\mathbf{x}$ is a positive definite quadratic form.

j. If A is a 2×2 symmetric matrix with positive entries and $\det(A) > 0$, then A is positive definite.

k. If A is symmetric, and if the quadratic form $\mathbf{x}^T A\mathbf{x}$ has no cross product terms, then A must be a diagonal matrix.

l. If $\mathbf{x}^T A\mathbf{x}$ is a positive definite quadratic form in two variables and $c \neq 0$, then the graph of the equation $\mathbf{x}^T A\mathbf{x} = c$ is an ellipse.

Working with Technology

T1. Find an orthogonal matrix P such that $P^T AP$ is diagonal.

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

T2. Use the eigenvalues of the following matrix to determine whether it is positive definite, negative definite, or indefinite, and then confirm your conclusion using Theorem 7.3.4.

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

7.4 Optimization Using Quadratic Forms

Quadratic forms arise in various problems in which the maximum or minimum value of some quantity is required. In this section we will discuss some problems of this type.

Constrained Extremum Problems

Our first goal in this section is to consider the problem of finding the maximum and minimum values of a quadratic form $\mathbf{x}^T A\mathbf{x}$ subject to the constraint $\|\mathbf{x}\| = 1$. Problems of this type arise in a wide variety of applications.

To visualize this problem geometrically in the case where $\mathbf{x}^T A\mathbf{x}$ is a quadratic form on R^2, view $z = \mathbf{x}^T A\mathbf{x}$ as the equation of some surface in a rectangular xyz-coordinate system and view $\|\mathbf{x}\| = 1$ as the unit circle centered at the origin of the xy-plane. Geometrically, the problem of finding the maximum and minimum values of $\mathbf{x}^T A\mathbf{x}$ subject to the requirement $\|\mathbf{x}\| = 1$ amounts to finding the highest and lowest points on the intersection of the surface with the right circular cylinder determined by the circle (**Figure 7.4.1**).

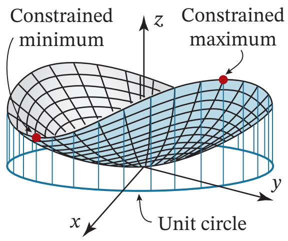


FIGURE 7.4.1

The following theorem, whose proof is deferred to the end of the section, is the key result for solving problems of this type.

Theorem 7.4.1

Constrained Extremum Theorem

Let A be a symmetric $n \times n$ matrix whose eigenvalues in order of decreasing size are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then:

(*a*) The quadratic form $\mathbf{x}^T A\mathbf{x}$ has a maximum value of λ_1 and a minimum value of λ_n, both of which are obtained on the set of vectors for which $\|\mathbf{x}\| = 1$.

(*b*) The maximum value of $\mathbf{x}^T A\mathbf{x}$ occurs at an eigenvector corresponding to the eigenvalue λ_1.

(*c*) The minimum value of $\mathbf{x}^T A\mathbf{x}$ occurs at an eigenvector corresponding to the eigenvalue λ_n.

Remark The condition $\|\mathbf{x}\| = 1$ in this theorem is called a ***constraint***, and the maximum or minimum value of $\mathbf{x}^T A\mathbf{x}$ subject to the constraint is called a ***constrained extremum***. This constraint can also be expressed as $\mathbf{x}^T\mathbf{x} = 1$ or as $x_1^2 + x_2^2 + \cdots + x_n^2 = 1$, when convenient.

EXAMPLE 1 | Finding Constrained Extrema

Find the maximum and minimum values of the quadratic form

$$z = 5x^2 + 5y^2 + 4xy$$

subject to the constraint $x^2 + y^2 = 1$.

Solution The quadratic form can be expressed in matrix notation as

$$z = 5x^2 + 5y^2 + 4xy = \mathbf{x}^T A\mathbf{x} = [x \quad y]\begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

We leave it for you to show that the eigenvalues of A are $\lambda_1 = 7$ and $\lambda_2 = 3$ and that corresponding eigenvectors are

$$\lambda_1 = 7: \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \lambda_2 = 3: \quad \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Normalizing these eigenvectors yields

$$\lambda_1 = 7: \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \lambda_2 = 3: \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \tag{1}$$

Thus, the constrained extrema are

$$\text{constrained maximum: } z = 7 \text{ at } (x, y) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

$$\text{constrained minimum: } z = 3 \text{ at } (x, y) = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

Remark Since the negatives of the eigenvectors in (1) are also unit eigenvectors, they too produce the maximum and minimum values of z; that is, the constrained maximum $z = 7$ also occurs at the point $(x, y) = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ and the constrained minimum $z = 3$ at $(x, y) = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$.

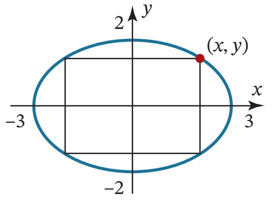


FIGURE 7.4.2 A rectangle inscribed in the ellipse $4x^2 + 9y^2 = 36$.

EXAMPLE 2 | A Constrained Extremum Problem

A rectangle is to be inscribed in the ellipse $4x^2 + 9y^2 = 36$, as shown in **Figure 7.4.2**. Use eigenvalue methods to find nonnegative values of x and y that produce the inscribed rectangle with maximum area.

Solution The area z of the inscribed rectangle is given by $z = 4xy$, so the problem is to maximize the quadratic form $z = 4xy$ subject to the constraint $4x^2 + 9y^2 = 36$. In this problem, the graph of the constraint equation is an ellipse rather than the unit circle as required in Theorem 7.4.1, but we can remedy this problem by rewriting the constraint as

$$\left(\frac{x}{3}\right)^2 + \left(\frac{y}{2}\right)^2 = 1$$

and defining new variables, x_1 and y_1, by the equations

$$x = 3x_1 \quad \text{and} \quad y = 2y_1$$

This enables us to reformulate the problem as follows:

$$\text{maximize } z = 4xy = 24x_1y_1$$

subject to the constraint

$$x_1^2 + y_1^2 = 1$$

To solve this problem, we will write the quadratic form $z = 24x_1y_1$ as

$$z = \mathbf{x}^T A\mathbf{x} = [x_1 \quad y_1]\begin{bmatrix} 0 & 12 \\ 12 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

We now leave it for you to show that the largest eigenvalue of A is $\lambda = 12$ and that the only corresponding unit eigenvector with nonnegative entries is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Thus, the maximum area is $z = 12$, and this occurs when

$$x = 3x_1 = \frac{3}{\sqrt{2}} \quad \text{and} \quad y = 2y_1 = \frac{2}{\sqrt{2}}$$

Constrained Extrema and Level Curves

A useful way of visualizing the behavior of a function $f(x, y)$ of two variables is to consider the curves in the xy-plane along which $f(x, y)$ is constant. These curves have equations of the form

$$f(x, y) = k$$

and are called the ***level curves*** of f (**Figure 7.4.3**). In particular, the level curves of a quadratic form $\mathbf{x}^T A\mathbf{x}$ on R^2 have equations of the form

$$\mathbf{x}^T A\mathbf{x} = k \tag{2}$$

so the maximum and minimum values of $\mathbf{x}^T A\mathbf{x}$ subject to the constraint $\|\mathbf{x}\| = 1$ are the largest and smallest values of k for which the graph of (2) intersects the unit circle. Typically, such values of k produce level curves that just touch the unit circle (**Figure 7.4.4**), and the coordinates of the points where the level curves just touch produce the vectors that maximize or minimize $\mathbf{x}^T A\mathbf{x}$ subject to the constraint $\|\mathbf{x}\| = 1$.

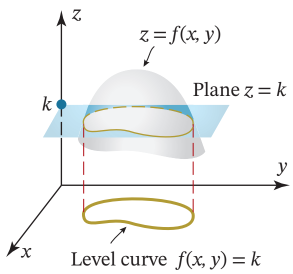


FIGURE 7.4.3

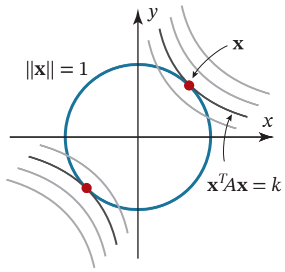


FIGURE 7.4.4

EXAMPLE 3 | Example 1 Revisited Using Level Curves

In Example 1 (and its following remark) we found the maximum and minimum values of the quadratic form

$$z = 5x^2 + 5y^2 + 4xy$$

subject to the constraint $x^2 + y^2 = 1$. We showed that the constrained maximum is $z = 7$, which is attained at the points

$$(x, y) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \quad \text{and} \quad (x, y) = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \tag{3}$$

and that the constrained minimum is $z = 3$, which is attained at the points

$$(x, y) = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \quad \text{and} \quad (x, y) = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \tag{4}$$

Geometrically, this means that the level curve $5x^2 + 5y^2 + 4xy = 7$ should just touch the unit circle at the points in (3), and the level curve $5x^2 + 5y^2 + 4xy = 3$ should just touch it at the points in (4). All of this is consistent with **Figure 7.4.5**.

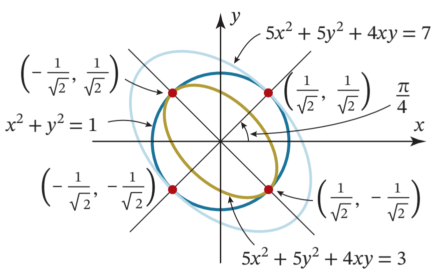


FIGURE 7.4.5

Relative Extrema of Functions of Two Variables

CALCULUS REQUIRED

We will conclude this section by showing how quadratic forms can be used to study characteristics of real-valued functions of two variables.

Recall that if a function $f(x, y)$ has first-order partial derivatives, then its relative maxima and minima, if any, occur at points where the conditions

$$f_x(x, y) = 0 \quad \text{and} \quad f_y(x, y) = 0$$

are both true. These are called ***critical points*** of f. The specific behavior of f at a critical point (x_0, y_0) is determined by the sign of

$$D(x, y) = f(x, y) - f(x_0, y_0) \tag{5}$$

at points (x, y) that are close to, but different from, (x_0, y_0):

- If $D(x, y) > 0$ at points (x, y) that are sufficiently close to, but different from, (x_0, y_0), then $f(x_0, y_0) < f(x, y)$ at such points and f is said to have a ***relative minimum*** at (x_0, y_0) (**Figure 7.4.6*a***).
- If $D(x, y) < 0$ at points (x, y) that are sufficiently close to, but different from, (x_0, y_0), then $f(x_0, y_0) > f(x, y)$ at such points and f is said to have a ***relative maximum*** at (x_0, y_0) (**Figure 7.4.6*b***).
- If $D(x, y)$ has both positive and negative values inside *every* circle centered at (x_0, y_0), then there are points (x, y) that are arbitrarily close to the point (x_0, y_0) at which $f(x_0, y_0) < f(x, y)$ and points (x, y) that are arbitrarily close to (x_0, y_0) at which $f(x_0, y_0) > f(x, y)$. In this case we say that f has a ***saddle point*** at (x_0, y_0) (**Figure 7.4.6*c***).

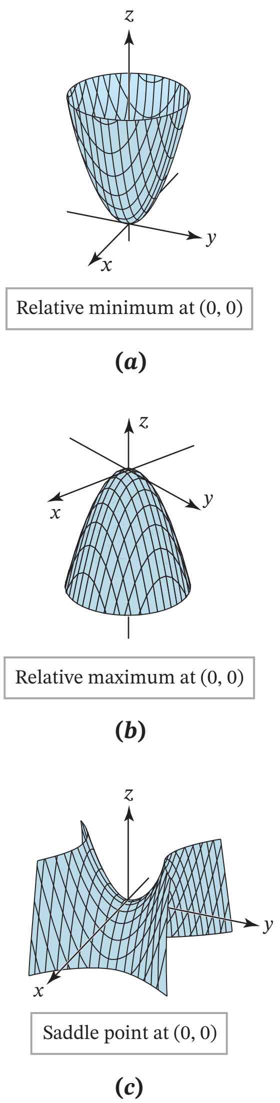


FIGURE 7.4.6

In general, it can be difficult to determine the sign of (5) directly. However, the following theorem, which is proved in calculus, makes it possible to analyze critical points using derivatives.

Theorem 7.4.2

Second Derivative Test

Suppose that (x_0, y_0) is a critical point of $f(x, y)$ and that f has continuous second-order partial derivatives in some circular region centered at (x_0, y_0). Then:

(a) f has a relative minimum at (x_0, y_0) if

$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) > 0 \quad \text{and} \quad f_{xx}(x_0, y_0) > 0$$

(b) f has a relative maximum at (x_0, y_0) if

$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) > 0 \quad \text{and} \quad f_{xx}(x_0, y_0) < 0$$

(c) f has a saddle point at (x_0, y_0) if

$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) < 0$$

(d) The test is inconclusive if

$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) = 0$$

Our interest here is in showing how to reformulate this theorem using properties of symmetric matrices. For this purpose we consider the symmetric matrix

$$H(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{xy}(x, y) & f_{yy}(x, y) \end{bmatrix}$$

which is called the ***Hessian*** or ***Hessian matrix*** of f in honor of the German mathematician and scientist Ludwig Otto Hesse (1811–1874). The notation $H(x, y)$ emphasizes that the entries in the matrix depend on x and y. The Hessian is of interest because

$$\det[H(x_0, y_0)] = \begin{vmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{xy}(x_0, y_0) & f_{yy}(x_0, y_0) \end{vmatrix} = f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0)$$

is the expression that appears in Theorem 7.4.2. We can now reformulate the second derivative test as follows.

Theorem 7.4.3

Hessian Form of the Second Derivative Test

Suppose that (x_0, y_0) is a critical point of $f(x, y)$ and that f has continuous second-order partial derivatives in some circular region centered at (x_0, y_0). If $H(x_0, y_0)$ is the Hessian of f at (x_0, y_0), then:

(a) f has a relative minimum at (x_0, y_0) if $H(x_0, y_0)$ is positive definite.

(b) f has a relative maximum at (x_0, y_0) if $H(x_0, y_0)$ is negative definite.

(c) f has a saddle point at (x_0, y_0) if $H(x_0, y_0)$ is indefinite.

(d) The test is inconclusive otherwise.

We will prove part (a). The proofs of the remaining parts will be left as exercises.

Proof (a) If $H(x_0, y_0)$ is positive definite, then Theorem 7.3.4 implies that the principal submatrices of $H(x_0, y_0)$ have positive determinants. Thus,

$$\det[H(x_0, y_0)] = \begin{vmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{xy}(x_0, y_0) & f_{yy}(x_0, y_0) \end{vmatrix} = f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) > 0$$

and

$$\det[f_{xx}(x_0, y_0)] = f_{xx}(x_0, y_0) > 0$$

so f has a relative minimum at (x_0, y_0) by part (a) of Theorem 7.4.2. ■

EXAMPLE 4 | Using the Hessian to Classify Relative Extrema

Find the critical points of the function

$$f(x, y) = \tfrac{1}{3}x^3 + xy^2 - 8xy + 3$$

and use the eigenvalues of the Hessian matrix at those points to determine which of them, if any, are relative maxima, relative minima, or saddle points.

Solution To find both the critical points and the Hessian matrix we will need to calculate the first and second partial derivatives of f. These derivatives are

$$\begin{aligned} &f_x(x, y) = x^2 + y^2 - 8y, \quad && f_y(x, y) = 2xy - 8x, \quad && f_{xy}(x, y) = 2y - 8 \\ &f_{xx}(x, y) = 2x, && f_{yy}(x, y) = 2x \end{aligned}$$

Thus, the Hessian matrix is

$$H(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{xy}(x, y) & f_{yy}(x, y) \end{bmatrix} = \begin{bmatrix} 2x & 2y - 8 \\ 2y - 8 & 2x \end{bmatrix}$$

To find the critical points we set f_x and f_y equal to zero. This yields the equations

$$f_x(x, y) = x^2 + y^2 - 8y = 0 \quad \text{and} \quad f_y(x, y) = 2xy - 8x = 2x(y - 4) = 0$$

Solving the second equation yields $x = 0$ or $y = 4$. Substituting $x = 0$ in the first equation and solving for y yields $y = 0$ or $y = 8$; and substituting $y = 4$ into the first equation and solving for x yields $x = 4$ or $x = -4$. Thus, we have four critical points:

$$(0, 0), \quad (0, 8), \quad (4, 4), \quad (-4, 4)$$

Evaluating the Hessian matrix at these points yields

$$H(0,0) = \begin{bmatrix} 0 & -8 \\ -8 & 0 \end{bmatrix}, \quad H(0,8) = \begin{bmatrix} 0 & 8 \\ 8 & 0 \end{bmatrix}$$

$$H(4,4) = \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix}, \quad H(-4,4) = \begin{bmatrix} -8 & 0 \\ 0 & -8 \end{bmatrix}$$

We leave it for you to find the eigenvalues of these matrices and deduce the following classifications of the stationary points:

Critical Point (x_0, y_0)	λ_1	λ_2	**Classification**
(0, 0)	8	−8	Saddle point
(0, 8)	8	−8	Saddle point
(4, 4)	8	8	Relative minimum
(−4, 4)	−8	−8	Relative maximum

OPTIONAL: We conclude this section with an optional proof of Theorem 7.4.1.

Proof of Theorem 7.4.1 The first step in the proof is to show that $\mathbf{x}^T A\mathbf{x}$ has constrained maximum and minimum values for $\|\mathbf{x}\| = 1$. Since A is symmetric, the principal axes theorem (Theorem 7.3.1) implies that there is an orthogonal change of variable $\mathbf{x} = P\mathbf{y}$ such that

$$\mathbf{x}^T A\mathbf{x} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \tag{6}$$

in which $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A. Let us assume that $\|\mathbf{x}\| = 1$ and that the column vectors of P (which are unit eigenvectors of A) have been ordered so that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \tag{7}$$

Since the matrix P is orthogonal, multiplication by P is length preserving, from which it follows that $\|\mathbf{y}\| = \|\mathbf{x}\| = 1$; that is,

$$y_1^2 + y_2^2 + \cdots + y_n^2 = 1$$

It follows from this equation and (7) that

$$\begin{aligned} \lambda_n = \lambda_n(y_1^2 + y_2^2 + \cdots + y_n^2) &\leq \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \\ &\leq \lambda_1(y_1^2 + y_2^2 + \cdots + y_n^2) = \lambda_1 \end{aligned}$$

and hence from (6) that

$$\lambda_n \leq \mathbf{x}^T A\mathbf{x} \leq \lambda_1$$

This shows that all values of $\mathbf{x}^T A\mathbf{x}$ for which $\|\mathbf{x}\| = 1$ lie between the largest and smallest eigenvalues of A. Now let $\mathbf{x}$ be a unit eigenvector corresponding to λ_1. Then

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T(\lambda_1\mathbf{x}) = \lambda_1\mathbf{x}^T\mathbf{x} = \lambda_1\|\mathbf{x}\|^2 = \lambda_1$$

which shows that $\mathbf{x}^T A\mathbf{x}$ has λ_1 as a constrained maximum and that this maximum occurs if $\mathbf{x}$ is a unit eigenvector of A corresponding to λ_1. Similarly, if $\mathbf{x}$ is a unit eigenvector corresponding to λ_n, then

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T(\lambda_n\mathbf{x}) = \lambda_n\mathbf{x}^T\mathbf{x} = \lambda_n\|\mathbf{x}\|^2 = \lambda_n$$

so $\mathbf{x}^T A\mathbf{x}$ has λ_n as a constrained minimum and this minimum occurs if $\mathbf{x}$ is a unit eigenvector of A corresponding to λ_n. This completes the proof. ■

Exercise Set 7.4

In Exercises 1–4, find the maximum and minimum values of the given quadratic form subject to the constraint $x^2 + y^2 = 1$, and determine the values of x and y at which the maximum and minimum occur.

1. $5x^2 - y^2$ **2.** xy **3.** $3x^2 + 7y^2$ **4.** $5x^2 + 5xy$

In Exercises 5–6, find the maximum and minimum values of the given quadratic form subject to the constraint

$$x^2 + y^2 + z^2 = 1$$

and determine the values of x, y, and z at which the maximum and minimum occur.

5. $9x^2 + 4y^2 + 3z^2$ **6.** $2x^2 + y^2 + z^2 + 2xy + 2xz$

7. Use the method of Example 2 to find the maximum and minimum values of xy subject to the constraint $4x^2 + 8y^2 = 16$.

8. Use the method of Example 2 to find the maximum and minimum values of $x^2 + xy + 2y^2$ subject to the constraint

$$x^2 + 3y^2 = 16$$

In Exercises 9–10, draw the unit circle and the level curves corresponding to the given quadratic form. Show that the unit circle intersects each of these curves in exactly two places, label the intersection points, and verify that the constrained extrema occur at those points.

9. $5x^2 - y^2$ **10.** xy

11. a. Show that the function $f(x, y) = 4xy - x^4 - y^4$ has critical points at $(0, 0)$, $(1, 1)$, and $(-1, -1)$.

b. Use the Hessian form of the second derivative test to show that f has relative maxima at $(1, 1)$ and $(-1, -1)$ and a saddle point at $(0, 0)$.

12. a. Show that the function $f(x, y) = x^3 - 6xy - y^3$ has critical points at $(0, 0)$ and $(-2, 2)$.

b. Use the Hessian form of the second derivative test to show that f has a relative maximum at $(-2, 2)$ and a saddle point at $(0, 0)$.

In Exercises 13–16, find the critical points of f, if any, and classify them as relative maxima, relative minima, or saddle points.

13. $f(x, y) = x^3 - 3xy - y^3$

14. $f(x, y) = x^3 - 3xy + y^3$

15. $f(x, y) = x^2 + 2y^2 - x^2y$

16. $f(x, y) = x^3 + y^3 - 3x - 3y$

17. A rectangle whose center is at the origin and whose sides are parallel to the coordinate axes is to be inscribed in the ellipse $x^2 + 25y^2 = 25$. Use the method of Example 2 to find nonnegative values of x and y that produce the inscribed rectangle with maximum area.

18. Suppose that $\mathbf{x}$ is a unit eigenvector of a matrix A corresponding to an eigenvalue 2. What is the value of $\mathbf{x}^T A\mathbf{x}$?

19. a. Show that the functions

$$f(x, y) = x^4 + y^4 \quad \text{and} \quad g(x, y) = x^4 - y^4$$

have a critical point at $(0, 0)$ but the second derivative test is inconclusive at that point.

b. Give a reasonable argument to show that f has a relative minimum at $(0, 0)$ and g has a saddle point at $(0, 0)$.

20. Suppose that the Hessian matrix of a certain quadratic form $f(x, y)$ is

$$H = \begin{bmatrix} 2 & 4 \\ 4 & 2 \end{bmatrix}$$

What can you say about the location and classification of the critical points of f?

21. Suppose that A is an $n \times n$ symmetric matrix and

$$q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$$

where $\mathbf{x}$ is a vector in R^n that is expressed in column form. What can you say about the value of q if $\mathbf{x}$ is a unit eigenvector corresponding to an eigenvalue λ of A?

Working with Proofs

22. Prove: If $\mathbf{x}^T A\mathbf{x}$ is a quadratic form whose minimum and maximum values subject to the constraint $\|\mathbf{x}\| = 1$ are m and M, respectively, then for each number c in the interval

$$m \le c \le M$$

there is a unit vector $\mathbf{x}_c$ such that $\mathbf{x}_c^T A\mathbf{x}_c = c$. [*Hint:* In the case where $m < M$, let $\mathbf{u}_m$ and $\mathbf{u}_M$ be unit eigenvectors of A such that $\mathbf{u}_m^T A\mathbf{u}_m = m$ and $\mathbf{u}_M^T A\mathbf{u}_M = M$, and let

$$\mathbf{x}_c = \sqrt{\frac{M - c}{M - m}}\mathbf{u}_m + \sqrt{\frac{c - m}{M - m}}\mathbf{u}_M$$

Show that $\mathbf{x}_c^T A\mathbf{x}_c = c$.]

True-False Exercises

TF. In parts (a)–(e) determine whether the statement is true or false, and justify your answer.

a. A quadratic form must have either a maximum or minimum value.

b. The maximum value of a quadratic form $\mathbf{x}^T A\mathbf{x}$ subject to the constraint $\|\mathbf{x}\| = 1$ occurs at a unit eigenvector corresponding to the largest eigenvalue of A.

c. The Hessian matrix of a function f with continuous second-order partial derivatives is a symmetric matrix.

d. If (x_0, y_0) is a critical point of a function f and the Hessian of f at (x_0, y_0) is 0, then f has neither a relative maximum nor a relative minimum at (x_0, y_0).

e. If A is a symmetric matrix and $\det(A) < 0$, then the minimum of $\mathbf{x}^T A\mathbf{x}$ subject to the constraint $\|\mathbf{x}\| = 1$ is negative.

Working with Technology

T1. Find the maximum and minimum values of the following quadratic form subject to the stated constraint, and specify the points at which those values are attained.

$$w = 2x^2 + y^2 + z^2 + 2xy + 2xz; \quad x^2 + y^2 + z^2 = 1$$

T2. Suppose that the temperature at a point (x, y) on a metal plate is $T(x, y) = 4x^2 - 4xy + y^2$. An ant walking on the plate traverses a circle of radius 5 centered at the origin. What are the highest and lowest temperatures encountered by the ant?

T3. The accompanying figure shows the intersection of the surface $z = x^2 + 4y^2$ (called an *elliptic paraboloid*) and the surface $x^2 + y^2 = 1$ (called a *right circular cylinder*). Find the highest and lowest points on the curve of intersection.

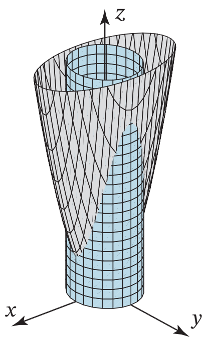


FIGURE Ex-T3

7.5 Hermitian, Unitary, and Normal Matrices

We showed in Section 7.2 that every symmetric matrix with real entries is orthogonally diagonalizable, and conversely that every diagonalizable matrix with real entries is symmetric. In this section we will be concerned with the diagonalization problem for matrices with complex entries.

Real Matrices Versus Complex Matrices

As discussed in Section 5.3, we distinguish between matrices whose entries must be real numbers, called ***real matrices***, and matrices whose entries may be either real numbers or complex numbers, called ***complex matrices***. When convenient, you can think of a real matrix as a complex matrix each of whose entries has zero as its imaginary part. Similarly, we distinguish between ***real vectors*** (those in R^n) and ***complex vectors*** (those in C^n).

Hermitian and Unitary Matrices

The transpose operation is less important for complex matrices than for real matrices. A more useful operation for complex matrices is given in the following definition.

Definition 1

If A is a complex matrix, then the ***conjugate transpose*** of A, denoted by A^*, is defined by

$$A^* = \overline{A}^T \tag{1}$$

Remark Note that the order in which the transpose and conjugation operations are performed in Formula (1) does not matter (see Theorem 5.3.2*b*). Moreover, if A is a real matrix, then Formula (1) simplifies to $A^* = (\overline{A})^T = A^T$, so the conjugate transpose is the same as the transpose in that case.

EXAMPLE 1 | Conjugate Transpose

Find the conjugate transpose A^* of the matrix

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

Solution We have

$$\overline{A} = \begin{bmatrix} 1-i & i & 0 \\ 2 & 3+2i & -i \end{bmatrix} \quad \text{and hence} \quad A^* = \overline{A}^T = \begin{bmatrix} 1-i & 2 \\ i & 3+2i \\ 0 & -i \end{bmatrix}$$

The following theorem, parts of which are given as exercises, shows that the basic algebraic properties of the conjugate transpose operation are similar to those of the transpose (compare to Theorem 1.4.8).

Theorem 7.5.1

If k is a complex scalar, and if A and B are complex matrices whose sizes are such that the stated operations can be performed, then:

(a) $(A^*)^* = A$

(b) $(A+B)^* = A^* + B^*$

(c) $(A-B)^* = A^* - B^*$

(d) $(kA)^* = \overline{k}A^*$

(e) $(AB)^* = B^*A^*$

We now define two new classes of matrices that will be important in our study of diagonalization in C^n.

Definition 2

A square matrix A is said to be ***unitary*** if

$$AA^* = A^*A = I \tag{2}$$

or, equivalently, if

$$A^* = A^{-1} \tag{3}$$

and it is said to be ***Hermitian**** if

$$A^* = A \tag{4}$$

To show that a matrix is unitary it suffices to show that *either* $AA^* = I$ or $A^*A = I$ since either equation implies the other.

If A is a real matrix, then $A^* = A^T$, in which case (3) becomes $A^T = A^{-1}$ and (4) becomes $A^T = A$. Thus, the unitary matrices are complex generalizations of the real orthogonal matrices and the Hermitian matrices are complex generalizations of the real symmetric matrices.

*In honor of the French mathematician Charles Hermite (1822–1901).

EXAMPLE 2 | Recognizing Hermitian Matrices

Hermitian matrices are easy to recognize because their diagonal entries are real (why?) and the entries that are symmetrically positioned across the main diagonal are complex conjugates. Thus, for example, we can tell by inspection that the following matrix is Hermitian:

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

EXAMPLE 3 | Recognizing Unitary Matrices

Unlike Hermitian matrices, unitary matrices are not readily identifiable by inspection. The most direct way to identify such matrices is to determine whether the matrix satisfies Equation (2) or Equation (3). We leave it for you to verify that the following matrix is unitary:

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}i \\ -\frac{1}{\sqrt{2}}i & \frac{1}{\sqrt{2}} \end{bmatrix}$$

In Theorem 7.2.2 we established that real symmetric matrices have real eigenvalues and that eigenvectors from different eigenvalues are orthogonal. That theorem is a special case of our next theorem in which orthogonality is with respect to the complex Euclidean inner product on C^n. We will prove part (b) of the theorem and leave the proof of part (a) for the exercises. In our proof we will make use of the fact that the relationship $\mathbf{u}\cdot\mathbf{v} = \overline{\mathbf{v}}^T\mathbf{u}$ given in Formula (5) of Section 5.3 can be expressed in terms of the conjugate transpose as

$$\boxed{\mathbf{u}\cdot\mathbf{v} = \mathbf{v}^*\mathbf{u}} \tag{5}$$

Theorem 7.5.2

If A is a Hermitian matrix, then:

(a) The eigenvalues of A are all real numbers.

(b) Eigenvectors from different eigenspaces are orthogonal.

Proof (b) Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors of A corresponding to distinct eigenvalues λ_1 and λ_2. Using Formula (5) and the facts that $\lambda_1 = \overline{\lambda_1}, \lambda_2 = \overline{\lambda_2}$, and $A = A^*$, we can write

$$\begin{aligned} \lambda_1(\mathbf{v}_2\cdot\mathbf{v}_1) = (\lambda_1\mathbf{v}_1)^*\mathbf{v}_2 = (A\mathbf{v}_1)^*\mathbf{v}_2 &= (\mathbf{v}_1^*A^*)\mathbf{v}_2 \\ = (\mathbf{v}_1^*A)\mathbf{v}_2 &= \mathbf{v}_1^*(A\mathbf{v}_2) \\ = \mathbf{v}_1^*(\lambda_2\mathbf{v}_2) &= \lambda_2(\mathbf{v}_1^*\mathbf{v}_2) = \lambda_2(\mathbf{v}_2\cdot\mathbf{v}_1) \end{aligned}$$

This implies that $(\lambda_1-\lambda_2)(\mathbf{v}_2\cdot\mathbf{v}_1) = 0$ and hence that $\mathbf{v}_2\cdot\mathbf{v}_1 = 0$ (since $\lambda_1 \neq \lambda_2$). ■

EXAMPLE 4 | Eigenvalues and Eigenvectors of a Hermitian Matrix

Confirm that the Hermitian matrix

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

has real eigenvalues and that eigenvectors from different eigenspaces are orthogonal.

Solution The characteristic polynomial of A is

$$\begin{aligned}\det(\lambda I - A) &= \begin{vmatrix} \lambda-2 & -1-i \\ -1+i & \lambda-3 \end{vmatrix} \\ &= (\lambda-2)(\lambda-3) - (-1-i)(-1+i) \\ &= (\lambda^2 - 5\lambda + 6) - 2 = (\lambda-1)(\lambda-4)\end{aligned}$$

so the eigenvalues of A are $\lambda = 1$ and $\lambda = 4$, which are real. Bases for the eigenspaces of A can be obtained by solving the linear system

$$\begin{bmatrix} \lambda-2 & -1-i \\ -1+i & \lambda-3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

with $\lambda = 1$ and with $\lambda = 4$. We leave it for you to do this and to show that the general solutions of these systems are

$$\lambda = 1: \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = t\begin{bmatrix} -1-i \\ 1 \end{bmatrix} \quad \text{and} \quad \lambda = 4: \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = t\begin{bmatrix} \frac{1}{2}(1+i) \\ 1 \end{bmatrix}$$

Thus, bases for these eigenspaces are

$$\lambda = 1: \quad \mathbf{v}_1 = \begin{bmatrix} -1-i \\ 1 \end{bmatrix} \quad \text{and} \quad \lambda = 4: \quad \mathbf{v}_2 = \begin{bmatrix} \frac{1}{2}(1+i) \\ 1 \end{bmatrix}$$

The vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal since

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = (-1-i)\left(\overline{\tfrac{1}{2}(1+i)}\right) + (1)(1) = \tfrac{1}{2}(-1-i)(1-i) + 1 = 0$$

and hence all scalar multiples of them are also orthogonal.

As noted in Example 3, unitary matrices are not easy to recognize by inspection. However, the following analog of Theorems 7.1.1 and 7.1.3, part of which is proved in the exercises, provides a way of ascertaining whether a matrix is unitary without computing its inverse.

Theorem 7.5.3

If A is an $n \times n$ matrix with complex entries, then the following are equivalent.

(*a*) A is unitary.

(*b*) $\|A\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ in C^n.

(*c*) $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}$ and $\mathbf{y}$ in C^n.

(*d*) The column vectors of A form an orthonormal set in C^n with respect to the complex Euclidean inner product.

(*e*) The row vectors of A form an orthonormal set in C^n with respect to the complex Euclidean inner product.

EXAMPLE 5 | A Unitary Matrix

Use Theorem 7.5.3 to show that

$$A = \begin{bmatrix} \frac{1}{2}(1+i) & \frac{1}{2}(1+i) \\ \frac{1}{2}(1-i) & \frac{1}{2}(-1+i) \end{bmatrix}$$

is unitary, and then find A^{-1}.

Solution We will show that the row vectors

$$\mathbf{r}_1 = \begin{bmatrix} \frac{1}{2}(1+i) & \frac{1}{2}(1+i) \end{bmatrix} \quad \text{and} \quad \mathbf{r}_2 = \begin{bmatrix} \frac{1}{2}(1-i) & \frac{1}{2}(-1+i) \end{bmatrix}$$

are orthonormal. The relevant computations are

$$\|\mathbf{r}_1\| = \sqrt{\left|\tfrac{1}{2}(1+i)\right|^2 + \left|\tfrac{1}{2}(1+i)\right|^2} = \sqrt{\tfrac{1}{2} + \tfrac{1}{2}} = 1$$

$$\|\mathbf{r}_2\| = \sqrt{\left|\tfrac{1}{2}(1-i)\right|^2 + \left|\tfrac{1}{2}(-1+i)\right|^2} = \sqrt{\tfrac{1}{2} + \tfrac{1}{2}} = 1$$

$$\begin{aligned} \mathbf{r}_1 \cdot \mathbf{r}_2 &= \left(\tfrac{1}{2}(1+i)\right)\left(\overline{\tfrac{1}{2}(1-i)}\right) + \left(\tfrac{1}{2}(1+i)\right)\left(\overline{\tfrac{1}{2}(-1+i)}\right) \\ &= \left(\tfrac{1}{2}(1+i)\right)\left(\tfrac{1}{2}(1+i)\right) + \left(\tfrac{1}{2}(1+i)\right)\left(\tfrac{1}{2}(-1-i)\right) = \tfrac{1}{2}i - \tfrac{1}{2}i = 0 \end{aligned}$$

Since we now know that A is unitary, it follows that

$$A^{-1} = A^* = \begin{bmatrix} \frac{1}{2}(1-i) & \frac{1}{2}(1+i) \\ \frac{1}{2}(1-i) & \frac{1}{2}(-1-i) \end{bmatrix}$$

You can confirm the validity of this result by showing that $AA^* = A^*A = I$.

Unitary Diagonalizability

Since unitary matrices are the complex analogs of the real orthogonal matrices, the following definition is a natural generalization of orthogonal diagonalizability for real matrices.

Definition 3

A square complex matrix A is said to be ***unitarily diagonalizable*** if there is a unitary matrix P such that $P^*AP = D$ is a complex diagonal matrix. Any such matrix P is said to ***unitarily diagonalize*** A.

Recall that a real symmetric $n \times n$ matrix A has an orthonormal set of n eigenvectors. and is orthogonally diagonalized by any $n \times n$ matrix whose column vectors are an orthonormal set of eigenvectors of A. Here is the complex analog of that result.

Theorem 7.5.4

Every $n \times n$ Hermitian matrix A has an orthonormal set of n eigenvectors. Moreover, A is unitarily diagonalized by any $n \times n$ matrix P whose column vectors form an orthonormal set of eigenvectors of A.

The procedure for unitarily diagonalizing a Hermitian matrix A is exactly the same as that for orthogonally diagonalizing a symmetric matrix:

Unitarily Diagonalizing a Hermitian Matrix

Step 1. Find a basis for each eigenspace of A.

Step 2. Apply the Gram–Schmidt process to each of these bases to obtain orthonormal bases for the eigenspaces.

Step 3. Form the matrix P whose column vectors are the basis vectors obtained in Step 2. This will be a unitary matrix (Theorem 7.5.3) and will unitarily diagonalize A.

EXAMPLE 6 | Unitary Diagonalization of a Hermitian Matrix

Find a matrix P that unitarily diagonalizes the Hermitian matrix

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

Solution We showed in Example 4 that the eigenvalues of A are $\lambda = 1$ and $\lambda = 4$ and that bases for the corresponding eigenspaces are

$$\lambda = 1\colon\quad \mathbf{v}_1 = \begin{bmatrix} -1-i \\ 1 \end{bmatrix} \quad \text{and} \quad \lambda = 4\colon\quad \mathbf{v}_2 = \begin{bmatrix} \frac{1}{2}(1+i) \\ 1 \end{bmatrix}$$

Since each eigenspace has only one basis vector, the Gram–Schmidt process is simply a matter of normalizing these basis vectors. We leave it for you to show that

$$\mathbf{p}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{bmatrix} \frac{-1-i}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix} \quad \text{and} \quad \mathbf{p}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \begin{bmatrix} \frac{1+i}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}$$

Thus, A is unitarily diagonalized by the matrix

$$P = [\mathbf{p}_1 \quad \mathbf{p}_2] = \begin{bmatrix} \frac{-1-i}{\sqrt{3}} & \frac{1+i}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{bmatrix}$$

Although it is a little tedious, you may want to check this result by showing that

$$P^*AP = \begin{bmatrix} \frac{-1+i}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1-i}{\sqrt{6}} & \frac{2}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix} \begin{bmatrix} \frac{-1-i}{\sqrt{3}} & \frac{1+i}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

Skew-Symmetric and Skew-Hermitian Matrices

We will now consider two more classes of matrices that play a role in the analysis of the diagonalization problem. A square *real* matrix A is said to be ***skew-symmetric*** if $A^T = -A$, and a square *complex* matrix A is said to be ***skew-Hermitian*** if $A^* = -A$. We leave it as an exercise to show that a skew-symmetric matrix must have zeros on the main diagonal, and a skew-Hermitian matrix must have zeros or pure imaginary numbers on the main diagonal. Here are two examples:

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

[skew-symmetric] **[skew-Hermitian]**

Normal Matrices

Hermitian matrices enjoy many, but not all, of the properties of real symmetric matrices. For example, we know that real symmetric matrices are orthogonally diagonalizable, and Hermitian matrices are unitarily diagonalizable. However, whereas the real symmetric matrices are the only orthogonally diagonalizable matrices, the Hermitian matrices do not constitute the entire class of unitarily diagonalizable complex matrices. Specifically, it can be proved that a square complex matrix A is unitarily diagonalizable if and only if

$$AA^* = A^*A \tag{6}$$

Matrices with this property are said to be ***normal***. Normal matrices include the Hermitian, skew-Hermitian, and unitary matrices in the complex case and the symmetric, skew-symmetric, and orthogonal matrices in the real case. The nonzero skew-symmetric matrices are particularly interesting because they are examples of real matrices that are not orthogonally diagonalizable but are unitarily diagonalizable.

A Comparison of Eigenvalues

We have seen that Hermitian matrices have real eigenvalues. In the exercises we will ask you to show that the eigenvalues of a skew-Hermitian matrix are either zero or purely imaginary (have real part of zero) and that the eigenvalues of unitary matrices have modulus 1. These ideas are illustrated schematically in **Figure 7.5.1**.

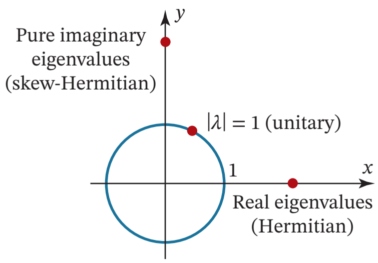


FIGURE 7.5.1

Exercise Set 7.5

In Exercises 1–2, find A^*.

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

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

In Exercises 3–4, substitute numbers for the ×'s so that A is Hermitian.

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

4. $A = \begin{bmatrix} 2 & 0 & 3+5i \\ \times & -4 & -i \\ \times & \times & 6 \end{bmatrix}$

In Exercises 5–6, show that A is not Hermitian for any choice of the ×'s.

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

b. $A = \begin{bmatrix} \times & \times & 3+5i \\ 0 & i & -i \\ 3-5i & i & \times \end{bmatrix}$

6. **a.** $A = \begin{bmatrix} 1 & 1+i & \times \\ 1+i & 7 & \times \\ 6-2i & \times & 0 \end{bmatrix}$

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

In Exercises 7–8, verify that the eigenvalues of the Hermitian matrix A are real and that eigenvectors from different eigenspaces are orthogonal (see Theorem 7.5.2).

7. $A = \begin{bmatrix} 3 & 2-3i \\ 2+3i & -1 \end{bmatrix}$ **8.** $A = \begin{bmatrix} 0 & 2i \\ -2i & 2 \end{bmatrix}$

In Exercises 9–12, show that A is unitary, and find A^{-1}.

9. $A = \begin{bmatrix} \frac{3}{5} & \frac{4}{5}i \\ -\frac{4}{5} & \frac{3}{5}i \end{bmatrix}$ **10.** $A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{2}(1+i) & \frac{1}{2}(1+i) \end{bmatrix}$

11. $A = \begin{bmatrix} \frac{1}{2\sqrt{2}}(\sqrt{3}+i) & \frac{1}{2\sqrt{2}}(1-i\sqrt{3}) \\ \frac{1}{2\sqrt{2}}(1+i\sqrt{3}) & \frac{1}{2\sqrt{2}}(i-\sqrt{3}) \end{bmatrix}$

12. $A = \begin{bmatrix} \frac{1}{\sqrt{3}}(-1+i) & \frac{1}{\sqrt{6}}(1-i) \\ \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{bmatrix}$

In Exercises 13–18, find a unitary matrix P that diagonalizes the Hermitian matrix A, and determine $P^{-1}AP$.

13. $A = \begin{bmatrix} 4 & 1-i \\ 1+i & 5 \end{bmatrix}$ **14.** $A = \begin{bmatrix} 3 & -i \\ i & 3 \end{bmatrix}$

15. $A = \begin{bmatrix} 6 & 2+2i \\ 2-2i & 4 \end{bmatrix}$ **16.** $A = \begin{bmatrix} 0 & 3+i \\ 3-i & -3 \end{bmatrix}$

17. $A = \begin{bmatrix} 5 & 0 & 0 \\ 0 & -1 & -1+i \\ 0 & -1-i & 0 \end{bmatrix}$

18. $A = \begin{bmatrix} 2 & \frac{1}{\sqrt{2}}i & -\frac{1}{\sqrt{2}}i \\ -\frac{1}{\sqrt{2}}i & 2 & 0 \\ \frac{1}{\sqrt{2}}i & 0 & 2 \end{bmatrix}$

In Exercises 19–20, substitute numbers for the $\times$'s so that A is skew-Hermitian.

19. $A = \begin{bmatrix} 0 & i & 2-3i \\ \times & 0 & 1 \\ \times & \times & 4i \end{bmatrix}$ **20.** $A = \begin{bmatrix} 0 & 0 & 3-5i \\ \times & 0 & -i \\ \times & \times & 0 \end{bmatrix}$

In Exercises 21–22, show that A is not skew-Hermitian for any choice of the $\times$'s.

21. **a.** $A = \begin{bmatrix} 0 & i & 2-3i \\ -i & 0 & \times \\ 2+3i & \times & \times \end{bmatrix}$

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

22. **a.** $A = \begin{bmatrix} i & \times & 2-3i \\ \times & 0 & 1+i \\ 2+3i & -1-i & \times \end{bmatrix}$

b. $A = \begin{bmatrix} 0 & -i & 4+7i \\ \times & 0 & \times \\ -4-7i & \times & 1 \end{bmatrix}$

In Exercises 23–24, verify that the eigenvalues of the skew-Hermitian matrix A are pure imaginary numbers.

23. $A = \begin{bmatrix} 0 & -1+i \\ 1+i & i \end{bmatrix}$ **24.** $A = \begin{bmatrix} 0 & 3i \\ 3i & 0 \end{bmatrix}$

In Exercises 25–26, show that A is normal.

25. $A = \begin{bmatrix} 1+2i & 2+i & -2-i \\ 2+i & 1+i & -i \\ -2-i & -i & 1+i \end{bmatrix}$

26. $A = \begin{bmatrix} 2+2i & i & 1-i \\ i & -2i & 1-3i \\ 1-i & 1-3i & -3+8i \end{bmatrix}$

27. Let A be any $n \times n$ matrix with complex entries, and define the matrices B and C to be

$$B = \tfrac{1}{2}(A + A^*) \quad \text{and} \quad C = \tfrac{1}{2i}(A - A^*)$$

a. Show that B and C are Hermitian.

b. Show that $A = B + iC$ and $A^* = B - iC$.

c. What condition must B and C satisfy for A to be normal?

28. Show that if A is an $n \times n$ matrix with complex entries, and if $\mathbf{u}$ and $\mathbf{v}$ are vectors in C^n that are expressed in column form, then

$$A\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot A^*\mathbf{v} \quad \text{and} \quad \mathbf{u} \cdot A\mathbf{v} = A^*\mathbf{u} \cdot \mathbf{v}$$

29. Show that

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{i\theta} & e^{-i\theta} \\ ie^{i\theta} & -ie^{-i\theta} \end{bmatrix}$$

is unitary for all real values of θ. [*Note:* See Formula (17) in Appendix B for the definition of $e^{i\theta}$.]

30. Show that

$$A = \begin{bmatrix} \alpha + i\gamma & -\beta + i\delta \\ \beta + i\delta & \alpha - i\gamma \end{bmatrix}$$

is unitary if $\alpha^2 + \beta^2 + \lambda^2 + \delta^2 = 1$.

31. Let A be the unitary matrix in Exercise 9, and verify that the conclusions in parts (*b*) and (*c*) of Theorem 7.5.3 hold for the vectors $\mathbf{x} = (1+i, 2-i)$ and $\mathbf{y} = (1, 1-i)$.

32. Let $T_A\colon C^2 \to C^2$ be multiplication by the Hermitian matrix A in Exercise 14, and find two orthogonal unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ for which $T_A(\mathbf{u}_1)$ and $T_A(\mathbf{u}_2)$ are orthogonal.

33. Under what conditions is the following matrix normal?

$$A = \begin{bmatrix} a & 0 & 0 \\ 0 & 0 & c \\ 0 & b & 0 \end{bmatrix}$$

34. What relationship must exist between a matrix and its inverse if it is both Hermitian and unitary?

35. Find a 2×2 matrix that is both Hermitian and unitary and whose entries are not all real numbers.

Working with Proofs

36. Use properties of the transpose and complex conjugate to prove parts (b) and (d) of Theorem 7.5.1.

37. Use properties of the transpose and complex conjugate to prove parts (a) and (e) of Theorem 7.5.1.

38. Prove that each entry on the main diagonal of a skew-Hermitian matrix is either zero or a pure imaginary number.

39. Prove that if A is a unitary matrix, then so is A^*.

40. Prove that the eigenvalues of a skew-Hermitian matrix are either zero or pure imaginary.

41. Prove that the eigenvalues of a unitary matrix have modulus 1.

42. Prove that if $\mathbf{u}$ is a nonzero vector in C^n that is expressed in column form, then $P = \mathbf{u}\mathbf{u}^*$ is Hermitian.

43. Prove that if $\mathbf{u}$ is a unit vector in C^n that is expressed in column form, then $H = I - 2\mathbf{u}\mathbf{u}^*$ is Hermitian and unitary.

44. Prove that if A is an invertible matrix, then A^* is invertible, and $(A^*)^{-1} = (A^{-1})^*$.

45. **a.** Prove that $\det(\overline{A}) = \overline{\det(A)}$.

b. Use the result in part (a) and the fact that a square matrix and its transpose have the same determinant to prove that $\det(A^*) = \overline{\det(A)}$.

46. Use part (b) of Exercise 45 to prove:

a. If A is Hermitian, then $\det(A)$ is real.

b. If A is unitary, then $|\det(A)| = 1$.

47. Prove that an $n \times n$ matrix with complex entries is unitary if and only if the columns of A form an orthonormal set in C^n.

48. Prove that the eigenvalues of a Hermitian matrix are real.

True-False Exercises

TF. In parts **(a)–(e)** determine whether the statement is true or false, and justify your answer.

a. The matrix $\begin{bmatrix} 0 & i \\ i & 2 \end{bmatrix}$ is Hermitian.

b. The matrix $\begin{bmatrix} -\frac{i}{\sqrt{2}} & \frac{i}{\sqrt{6}} & \frac{i}{\sqrt{3}} \\ 0 & -\frac{i}{\sqrt{6}} & \frac{i}{\sqrt{3}} \\ \frac{i}{\sqrt{2}} & \frac{i}{\sqrt{6}} & \frac{i}{\sqrt{3}} \end{bmatrix}$ is unitary.

c. The conjugate transpose of a unitary matrix is unitary.

d. Every unitarily diagonalizable matrix is Hermitian.

e. A positive integer power of a skew-Hermitian matrix is skew-Hermitian.

Chapter 7 Supplementary Exercises

1. Verify that each matrix is orthogonal, and find its inverse.

a. $\begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix}$ **b.** $\begin{bmatrix} \frac{4}{5} & 0 & -\frac{3}{5} \\ -\frac{9}{25} & \frac{4}{5} & -\frac{12}{25} \\ \frac{12}{25} & \frac{3}{5} & \frac{16}{25} \end{bmatrix}$

2. Prove: If Q is an orthogonal matrix, then each entry of Q is the same as its cofactor if $\det(Q) = 1$ and is the negative of its cofactor if $\det(Q) = -1$.

3. Prove that if A is a positive definite symmetric matrix, and if $\mathbf{u}$ and $\mathbf{v}$ are vectors in R^n in column form, then

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T A \mathbf{v}$$

is an inner product on R^n.

4. Find the characteristic polynomial and the dimensions of the eigenspaces of the symmetric matrix

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

5. Find a matrix P that orthogonally diagonalizes

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

and determine the diagonal matrix $D = P^T A P$.

6. Express each quadratic form in the matrix notation $\mathbf{x}^T A \mathbf{x}$.

a. $-4x_1^2 + 16x_2^2 - 15x_1x_2$

b. $9x_1^2 - x_2^2 + 4x_3^2 + 6x_1x_2 - 8x_1x_3 + x_2x_3$

7. Classify the quadratic form

$$x_1^2 - 3x_1x_2 + 4x_2^2$$

as positive definite, negative definite, indefinite, positive semidefinite, or negative semidefinite.

8. Find an orthogonal change of variable that eliminates the cross product terms in each quadratic form, and express the quadratic form in terms of the new variables.

a. $-3x_1^2 + 5x_2^2 + 2x_1x_2$

b. $-5x_1^2 + x_2^2 - x_3^2 + 6x_1x_3 + 4x_1x_2$

9. Identify the type of conic section represented by each equation.

a. $y - x^2 = 0$ **b.** $3x - 11y^2 = 0$

10. Find a unitary matrix U that diagonalizes

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

and determine the diagonal matrix $D = U^{-1} A U$.

11. Show that if U is an $n \times n$ unitary matrix and

$$|z_1| = |z_2| = \cdots = |z_n| = 1$$

then the product

$$U\begin{bmatrix} z_1 & 0 & 0 & \cdots & 0 \\ 0 & z_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & z_n \end{bmatrix}$$

is also unitary.

12. Show that:

a. The matrix iA is skew-Hermitian if and only if A is Hermitian.

b. If A is skew-Hermitian, then A is unitarily diagonalizable and has pure imaginary eigenvalues.

13. Find a, b, and c for which the matrix

$$\begin{bmatrix} a & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ b & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ c & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

is orthogonal. Are the values of a, b, and c unique? Explain.

14. In each part, suppose that A is a 4×4 matrix in which $\det(M_j)$ is the determinant of the jth principal submatrix of A. Determine whether A is positive definite, negative definite, or indefinite.

a. $\det(M_1) < 0, \det(M_2) > 0, \det(M_3) < 0, \det(M_4) > 0$

b. $\det(M_1) > 0, \det(M_2) > 0, \det(M_3) > 0, \det(M_4) > 0$

c. $\det(M_1) < 0, \det(M_2) < 0, \det(M_3) < 0, \det(M_4) < 0$

d. $\det(M_1) > 0, \det(M_2) < 0, \det(M_3) > 0, \det(M_4) < 0$

e. $\det(M_1) = 0, \det(M_2) < 0, \det(M_3) = 0, \det(M_4) > 0$

f. $\det(M_1) = 0, \det(M_2) > 0, \det(M_3) = 0, \det(M_4) = 0$

15. Prove:

a. If Q is an $m \times n$ matrix, then $C = QQ^T$ is symmetric and positive semidefinite.

b. The eigenvalues of C are nonnegative. [*Suggestion*: Look at the proof of Theorem 7.3.2 (a)].

CHAPTER **8**

General Linear Transformations

CHAPTER CONTENTS

Introduction

In earlier sections we studied linear transformations from R^n to R^m. In this chapter we will define and study linear transformations from a general vector space V to a general vector space W. The results we will obtain here have important applications in physics, engineering, and various branches of mathematics.

8.1 General Linear Transformations

Up to now our study of linear transformations has focused on transformations from R^n to R^m. In this section we will turn our attention to linear transformations involving general vector spaces. We will illustrate ways in which such transformations arise, and we will establish a fundamental relationship between general n-dimensional vector spaces and R^n.

Definitions and Terminology

In Section 1.8 we defined a *matrix transformation* $T_A: R^n \to R^m$ to be a mapping of the form

$$T_A(\mathbf{x}) = A\mathbf{x}$$

in which A is an $m \times n$ matrix. We subsequently established in Theorem 1.8.3 that the matrix transformations are precisely the *linear transformations* from R^n to R^m; that is, the transformations with the linearity properties

$$T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \quad \text{and} \quad T(k\mathbf{u}) = kT(\mathbf{u})$$

We will use these two properties as the starting point for defining more general linear transformations.

Definition 1

If $T: V \to W$ is a mapping from a vector space V to a vector space W, then T is called a ***linear transformation*** from V to W if the following two properties hold for all vectors $\mathbf{u}$ and $\mathbf{v}$ in V and for all scalars k:

(i) $T(k\mathbf{u}) = kT(\mathbf{u})$ **[Homogeneity property]**

(ii) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ **[Additivity property]**

In the special case where $V = W$, the linear transformation T is called a ***linear operator*** on the vector space V.

The homogeneity and additivity properties of a linear transformation $T: V \to W$ can be used in combination to show that if $\mathbf{v}_1$ and $\mathbf{v}_2$ are vectors in V and k_1 and k_2 are any scalars, then

$$T(k_1\mathbf{v}_1 + k_2\mathbf{v}_2) = k_1T(\mathbf{v}_1) + k_2T(\mathbf{v}_2)$$

More generally, if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are vectors in V and $k_1, k_2, \ldots, k_n$ are any scalars, then

$$\boxed{T(k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n) = k_1T(\mathbf{v}_1) + k_2T(\mathbf{v}_2) + \cdots + k_nT(\mathbf{v}_n)} \tag{1}$$

The following theorem is an analog of parts (a) and (d) of Theorem 1.8.1.

Theorem 8.1.1

If $T: V \to W$ is a linear transformation, then:

(a) $T(\mathbf{0}) = \mathbf{0}$.

(b) $T(\mathbf{u}-\mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$ for all $\mathbf{u}$ and $\mathbf{v}$ in V.

(c) $T(-\mathbf{v}) = -T(\mathbf{v})$ for all $\mathbf{v}$ in V.

Proof Let $\mathbf{u}$ be any vector in V. Since $0\mathbf{u} = \mathbf{0}$, it follows from the homogeneity property in Definition 1 that

$$T(\mathbf{0}) = T(0\mathbf{u}) = 0T(\mathbf{u}) = \mathbf{0}$$

which proves (a). We can prove part (b) by rewriting $T(\mathbf{u}-\mathbf{v})$ as

$$\begin{aligned} T(\mathbf{u}-\mathbf{v}) &= T\big(\mathbf{u} + (-1)\mathbf{v}\big) \\ &= T(\mathbf{u}) + (-1)T(\mathbf{v}) \\ &= T(\mathbf{u}) - T(\mathbf{v}) \end{aligned}$$

We leave it for you to justify each step. To prove part (c) set $\mathbf{u} = \mathbf{0}$ in part (b) and apply part (a). ■

EXAMPLE 1 | Matrix Transformations

Because we have based the definition of a general linear transformation on the homogeneity and additivity properties of *matrix transformations*, it follows that every matrix transformation $T_A: R^n \to R^m$ is a linear transformation in the sense of Definition 1.

EXAMPLE 2 | The Zero Transformation

Let V and W be any two vector spaces. The mapping $T: V \to W$ defined by $T(\mathbf{v}) = \mathbf{0}$ for every $\mathbf{v}$ in V is a linear transformation called the ***zero transformation***. To see that T is linear, observe that

$$T(\mathbf{u}+\mathbf{v}) = \mathbf{0}, \quad T(\mathbf{u}) = \mathbf{0}, \quad T(\mathbf{v}) = \mathbf{0}, \quad \text{and} \quad T(k\mathbf{u}) = \mathbf{0}$$

Therefore,

$$T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \quad \text{and} \quad T(k\mathbf{u}) = kT(\mathbf{u})$$

EXAMPLE 3 | The Identity Operator

Let V be any vector space. The mapping $I: V \to V$ defined by $I(\mathbf{v}) = \mathbf{v}$ is called the ***identity operator*** on V. We will leave it for you to verify that I is linear.

EXAMPLE 4 | Dilation and Contraction Operators

If V is a vector space and c is any scalar, then the linear operator $T: V \to V$ that is defined by $T(\mathbf{x}) = c\mathbf{x}$ is a linear operator on V, for if c is any scalar and if $\mathbf{u}$ and $\mathbf{v}$ are any vectors in V, then

$$T(k\mathbf{u}) = c(k\mathbf{u}) = k(c\mathbf{u}) = kT(\mathbf{u})$$
$$T(\mathbf{u}+\mathbf{v}) = c(\mathbf{u}+\mathbf{v}) = c\mathbf{u} + c\mathbf{v} = T(\mathbf{u}) + T(\mathbf{v})$$

If $0 < c < 1$, then T is called the ***contraction*** of V with factor c, and if $c > 1$, it is called the ***dilation*** of V with factor c.

EXAMPLE 5 | A Linear Transformation from P_n to P_{n+1}

Let $\mathbf{p} = p(x) = c_0 + c_1x + \cdots + c_nx^n$ be a polynomial in P_n, and define the transformation $T: P_n \to P_{n+1}$ by

$$T(\mathbf{p}) = T(p(x)) = xp(x) = c_0x + c_1x^2 + \cdots + c_nx^{n+1}$$

This transformation is linear because for any scalar k and any polynomials $\mathbf{p}_1$ and $\mathbf{p}_2$ in P_n we have

$$T(k\mathbf{p}) = T(kp(x)) = x(kp(x)) = k(xp(x)) = kT(\mathbf{p})$$

and

$$\begin{aligned} T(\mathbf{p}_1 + \mathbf{p}_2) &= T(p_1(x) + p_2(x)) = x(p_1(x) + p_2(x)) \\ &= xp_1(x) + xp_2(x) = T(\mathbf{p}_1) + T(\mathbf{p}_2) \end{aligned}$$

EXAMPLE 6 | A Linear Transformation Using an Inner Product

Let $\mathbf{v}_0$ be any fixed vector in a real inner product space V, and let $T: V \to R$ be the transformation

$$T(\mathbf{x}) = \langle \mathbf{x}, \mathbf{v}_0 \rangle$$

that maps a vector $\mathbf{x}$ to its inner product with $\mathbf{v}_0$. This transformation is linear, for if k is any scalar, and if $\mathbf{u}$ and $\mathbf{v}$ are any vectors in R^n, then it follows from properties of real inner products that

$$T(k\mathbf{u}) = \langle k\mathbf{u}, \mathbf{v}_0 \rangle = k\langle \mathbf{u}, \mathbf{v}_0 \rangle = kT(\mathbf{u})$$
$$T(\mathbf{u} + \mathbf{v}) = \langle \mathbf{u} + \mathbf{v}, \mathbf{v}_0 \rangle = \langle \mathbf{u}, \mathbf{v}_0 \rangle + \langle \mathbf{v}, \mathbf{v}_0 \rangle = T(\mathbf{u}) + T(\mathbf{v})$$

EXAMPLE 7 | Transformations on Matrix Spaces

Let M_{nn} be the vector space of $n \times n$ matrices. In each part determine whether the transformation is linear.

(*a*) $T_1(A) = A^T$ (*b*) $T_2(A) = \det(A)$

Solution (*a*) It follows from parts (*b*) and (*d*) of Theorem 1.4.8 that

$$T_1(kA) = (kA)^T = kA^T = kT_1(A)$$
$$T_1(A + B) = (A + B)^T = A^T + B^T = T_1(A) + T_1(B)$$

so T_1 is linear.

Solution (*b*) It follows from Formula (1) of Section 2.3 that

$$T_2(kA) = \det(kA) = k^n \det(A) = k^n T_2(A)$$

Thus, T_2 is not homogeneous and hence not linear if $n > 1$. Note that additivity also fails because we showed in Example 1 of Section 2.3 that $\det(A + B)$ and $\det(A) + \det(B)$ are not generally equal.

EXAMPLE 8 | Translation Is Not Linear

Part (*a*) of Theorem 8.1.1 states that a linear transformation maps $\mathbf{0}$ to $\mathbf{0}$. This property is useful for identifying transformations that are *not* linear. For example, if $\mathbf{x}_0$ is a fixed nonzero vector in a real inner product space V, then the transformation

$$T(\mathbf{x}) = \mathbf{x} + \mathbf{x}_0$$

has the geometric effect of translating each point $\mathbf{x}$ in a direction parallel to $\mathbf{x}_0$ through a distance of $\|\mathbf{x}_0\|$ (**Figure 8.1.1**). This cannot be a linear transformation since $T(\mathbf{0}) = \mathbf{x}_0$, so T does not map $\mathbf{0}$ to $\mathbf{0}$.

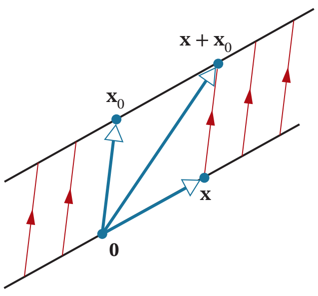


FIGURE 8.1.1 $T(\mathbf{x}) = \mathbf{x} + \mathbf{x}_0$ translates each point $\mathbf{x}$ along a line parallel to $\mathbf{x}_0$ through a distance $\|\mathbf{x}_0\|$.

EXAMPLE 9 | The Evaluation Transformation

Let V be a subspace of $F(-\infty, \infty)$, let

$$x_1, x_2, \ldots, x_n$$

be a sequence of distinct real numbers, and let $T: V \to R^n$ be the transformation

$$T(f) = \big(f(x_1), f(x_2), \ldots, f(x_n)\big) \tag{2}$$

that associates with the function f the n-tuple of function values at $x_1, x_2, \ldots, x_n$. We call this the ***evaluation transformation*** on V at $x_1, x_2, \ldots, x_n$. Thus, for example, if

$$x_1 = -1, \quad x_2 = 2, \quad x_3 = 4$$

and if $f(x) = x^2 - 1$, then

$$T(f) = \big(f(x_1), f(x_2), f(x_3)\big) = (0, 3, 15)$$

The evaluation transformation in (2) is linear, for if k is any scalar, and if f and g are any functions in V, then

$$\begin{aligned} T(kf) &= \big((kf)(x_1), (kf)(x_2), \ldots, (kf)(x_n)\big) \\ &= \big(kf(x_1), kf(x_2), \ldots, kf(x_n)\big) \\ &= k\big(f(x_1), f(x_2), \ldots, f(x_n)\big) = kT(f) \end{aligned}$$

and

$$\begin{aligned} T(f+g) &= \big((f+g)(x_1), (f+g)(x_2), \ldots, (f+g)(x_n)\big) \\ &= \big(f(x_1)+g(x_1), f(x_2)+g(x_2), \ldots, f(x_n)+g(x_n)\big) \\ &= \big(f(x_1), f(x_2), \ldots, f(x_n)\big) + \big(g(x_1), g(x_2), \ldots, g(x_n)\big) \\ &= T(f) + T(g) \end{aligned}$$

Finding Linear Transformations from Images of Basis Vectors

We saw in Formula (15) of Section 1.8 that if $T: R^n \to R^m$ is a linear transformation, and if $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ are the standard basis vectors for R^n, then the matrix A for T can be expressed as

$$A = [T(\mathbf{e}_1) \mid T(\mathbf{e}_2) \mid \cdots \mid T(\mathbf{e}_n)]$$

It follows from this that the image of any vector $\mathbf{v} = (c_1, c_2, \ldots, c_n)$ in R^n under multiplication by A can be expressed as

$$T(\mathbf{v}) = c_1T(\mathbf{e}_1) + c_2T(\mathbf{e}_2) + \cdots + c_nT(\mathbf{e}_n)$$

This formula tells us that for a matrix transformation the image of any vector is expressible as a linear combination of the images of the standard basis vectors. This is a special case of the following more general result.

Theorem 8.1.2

Let $T: V \to W$ be a linear transformation, for which the vector space V is finite-dimensional. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for V, then the image of any vector $\mathbf{v}$ in V can be expressed as

$$T(\mathbf{v}) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_nT(\mathbf{v}_n) \tag{3}$$

where $c_1, c_2, \ldots, c_n$ are the coefficients required to express $\mathbf{v}$ as a linear combination of the vectors in the basis S.

Proof Express $\mathbf{v}$ as $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$ and use the linearity of T. ■

EXAMPLE 10 | Computing with Images of Basis Vectors

Consider the basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for R^3, where

$$\mathbf{v}_1 = (1,1,1), \quad \mathbf{v}_2 = (1,1,0), \quad \mathbf{v}_3 = (1,0,0)$$

Let $T: R^3 \to R^2$ be the linear transformation for which

$$T(\mathbf{v}_1) = (1,0), \quad T(\mathbf{v}_2) = (2,-1), \quad T(\mathbf{v}_3) = (4,3)$$

Find a formula for $T(x_1, x_2, x_3)$, and then use that formula to compute $T(2,-3,5)$.

Solution We first need to express $\mathbf{x} = (x_1, x_2, x_3)$ as a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. If we write

$$(x_1, x_2, x_3) = c_1(1,1,1) + c_2(1,1,0) + c_3(1,0,0)$$

then on equating corresponding components, we obtain

$$\begin{aligned} c_1 + c_2 + c_3 &= x_1 \\ c_1 + c_2 \phantom{{}+c_3} &= x_2 \\ c_1 \phantom{{}+c_2+c_3} &= x_3 \end{aligned}$$

which yields $c_1 = x_3$, $c_2 = x_2 - x_3$, $c_3 = x_1 - x_2$, so

$$\begin{aligned} (x_1, x_2, x_3) &= x_3(1,1,1) + (x_2 - x_3)(1,1,0) + (x_1 - x_2)(1,0,0) \\ &= x_3\mathbf{v}_1 + (x_2 - x_3)\mathbf{v}_2 + (x_1 - x_2)\mathbf{v}_3 \end{aligned}$$

Thus

$$\begin{aligned} T(x_1, x_2, x_3) &= x_3 T(\mathbf{v}_1) + (x_2 - x_3)T(\mathbf{v}_2) + (x_1 - x_2)T(\mathbf{v}_3) \\ &= x_3(1,0) + (x_2 - x_3)(2,-1) + (x_1 - x_2)(4,3) \\ &= (4x_1 - 2x_2 - x_3, 3x_1 - 4x_2 + x_3) \end{aligned}$$

From this formula we obtain

$$T(2,-3,5) = (9,23)$$

EXAMPLE 11 | A Linear Transformation from $C^1(-\infty, \infty)$ to $F(-\infty, \infty)$

CALCULUS REQUIRED

Let $V = C^1(-\infty, \infty)$ be the vector space of functions with continuous first derivatives on $(-\infty, \infty)$, and let $W = F(-\infty, \infty)$ be the vector space of all real-valued functions defined on $(-\infty, \infty)$. Let $D: V \to W$ be the transformation that maps a function $\mathbf{f} = f(x)$ into its derivative—that is,

$$D(\mathbf{f}) = f'(x)$$

From the properties of differentiation, we have

$$D(\mathbf{f} + \mathbf{g}) = D(\mathbf{f}) + D(\mathbf{g}) \quad \text{and} \quad D(k\mathbf{f}) = kD(\mathbf{f})$$

Thus, D is a linear transformation.

CALCULUS REQUIRED

EXAMPLE 12 | An Integral Transformation

Let $V = C(-\infty, \infty)$ be the vector space of continuous functions on the interval $(-\infty, \infty)$, let $W = C^1(-\infty, \infty)$ be the vector space of functions with continuous first derivatives on $(-\infty, \infty)$, and let $J : V \to W$ be the transformation that maps a function f in V into

$$J(f) = \int_0^x f(t)\,dt$$

For example, if $f(x) = x^2$, then

$$J(f) = \int_0^x t^2\,dt = \frac{t^3}{3}\Big]_0^x = \frac{x^3}{3}$$

The transformation $J : V \to W$ is linear, for if k is any constant, and if f and g are any functions in V, then properties of the integral imply that

$$J(kf) = \int_0^x kf(t)\,dt = k\int_0^x f(t)\,dt = kJ(f)$$

$$J(f+g) = \int_0^x (f(t)+g(t))\,dt = \int_0^x f(t)\,dt + \int_0^x g(t)\,dt = J(f) + J(g)$$

Kernel and Range

Recall that if A is an $m \times n$ matrix, then the null space of A consists of all vectors $\mathbf{x}$ in R^n such that $A\mathbf{x} = \mathbf{0}$, and by Theorem 4.8.1 the column space of A consists of all vectors $\mathbf{b}$ in R^m for which there is at least one vector $\mathbf{x}$ in R^n such that $A\mathbf{x} = \mathbf{b}$. From the viewpoint of matrix transformations, the null space of A consists of all vectors in R^n that multiplication by A maps into $\mathbf{0}$, and the column space of A consists of all vectors in R^m that are images of at least one vector in R^n under multiplication by A. The following definition extends these ideas to general linear transformations, which is illustrated in (**Figure 8.1.2**).

Definition 2

If $T : V \to W$ is a linear transformation, then the set of vectors in V that T maps into $\mathbf{0}$ is called the ***kernel*** of T and is denoted by $\ker(T)$. The set of all vectors in W that are images under T of at least one vector in V is called the ***range*** of T and is denoted by $R(T)$.

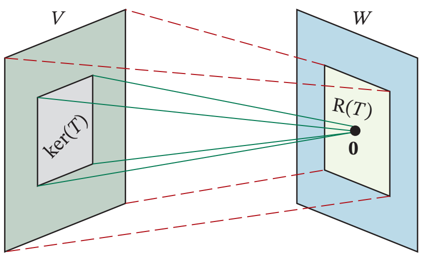


FIGURE 8.1.2

EXAMPLE 13 | Kernel and Range of a Matrix Transformation

If $T_A : R^n \to R^m$ is multiplication by the $m \times n$ matrix A, then the kernel of T_A is the null space of A, and the range of T_A is the column space of A.

EXAMPLE 14 | Kernel and Range of the Zero Transformation

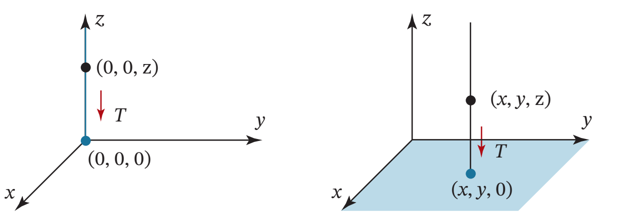
Let $T : V \to W$ be the zero transformation. Since T maps *every* vector in V into $\mathbf{0}$, it follows that $\ker(T) = V$. Moreover, since $\mathbf{0}$ is the *only* image under T of vectors in V, it follows that $R(T) = \{\mathbf{0}\}$.

EXAMPLE 15 | Kernel and Range of the Identity Operator

Let $I : V \to V$ be the identity operator. Since $I(\mathbf{v}) = \mathbf{v}$ for all vectors in V, *every* vector in V is the image of some vector (namely, itself); thus $R(I) = V$. Since the *only* vector that I maps into $\mathbf{0}$ is $\mathbf{0}$, it follows that $\ker(I) = \{\mathbf{0}\}$.

EXAMPLE 16 | Kernel and Range of an Orthogonal Projection

Let $T : R^3 \to R^3$ be the orthogonal projection onto the xy-plane. As illustrated in **Figure 8.1.3*a***, the points that T maps into $\mathbf{0} = (0, 0, 0)$ are precisely those on the z-axis, so $\ker(T)$ is the set of points of the form $(0, 0, z)$. As illustrated in **Figure 8.1.3*b***, T maps the points in R^3 to the xy-plane, where each point in that plane is the image of each point on the vertical line above it. Thus, $R(T)$ is the set of points of the form $(x, y, 0)$.



(*a*) $\ker(T)$ is the z-axis. (*b*) $R(T)$ is the entire xy-plane.

FIGURE 8.1.3

EXAMPLE 17 | Kernel and Range of a Rotation

Let $T : R^2 \to R^2$ be the linear operator that rotates each vector in the xy-plane through the angle θ (**Figure 8.1.4**). Since *every* vector in the xy-plane can be obtained by rotating some vector through the angle θ, it follows that $R(T) = R^2$. Moreover, the only vector that rotates into $\mathbf{0}$ is $\mathbf{0}$, so $\ker(T) = \{\mathbf{0}\}$.

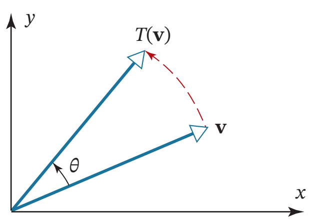


FIGURE 8.1.4

CALCULUS REQUIRED

EXAMPLE 18 | Kernel of a Differentiation Transformation

Let $V = C^1(-\infty,\infty)$ be the vector space of functions with continuous first derivatives on $(-\infty,\infty)$, let $W = F(-\infty,\infty)$ be the vector space of all real-valued functions defined on $(-\infty,\ \infty)$, and let $D: V \to W$ be the differentiation transformation $D(\mathbf{f}) = f'(x)$. The kernel of D is the set of functions in V with derivative zero. As shown in calculus, this is the set of constant functions on $(-\infty,\infty)$.

Properties of Kernel and Range

In all of the preceding examples, $\ker(T)$ and $R(T)$ turned out to be *subspaces*. In Examples 14, 15, and 17 they were either the zero subspace or the entire vector space. In Example 16 the kernel was a line through the origin, and the range was a plane through the origin, both of which are subspaces of R^3. All of this is a consequence of the following general theorem.

Theorem 8.1.3

If $T: V \to W$ is a linear transformation, then:

(a) The kernel of T is a subspace of V.

(b) The range of T is a subspace of W.

Proof (a) To show that $\ker(T)$ is a subspace, we must show that it contains at least one vector and is closed under addition and scalar multiplication. By part (a) of Theorem 8.1.1, the vector $\mathbf{0}$ is in $\ker(T)$, so the kernel contains at least one vector. If $\mathbf{0}$ is the only vector in the kernel of T, then $\ker(T)$ is the zero subspace of V. If there are at least two vectors in the kernel, then let $\mathbf{v}_1$ and $\mathbf{v}_2$ be any two such vectors, and let k be any scalar. Then

$$T(\mathbf{v}_1+\mathbf{v}_2) = T(\mathbf{v}_1)+T(\mathbf{v}_2) = \mathbf{0}+\mathbf{0} = \mathbf{0}$$

so $\mathbf{v}_1+\mathbf{v}_2$ is in $\ker(T)$. Also,

$$T(k\mathbf{v}_1) = kT(\mathbf{v}_1) = k\mathbf{0} = \mathbf{0}$$

so $k\mathbf{v}_1$ is in $\ker(T)$.

Proof (b) To show that $R(T)$ is a subspace of W, we must show that it contains at least one vector and is closed under addition and scalar multiplication. However, it contains at least the zero vector of W since $T(\mathbf{0}) = \mathbf{0}$ by part (a) of Theorem 8.1.1. To prove that it is closed under addition and scalar multiplication, we must show that if $\mathbf{w}_1$ and $\mathbf{w}_2$ are vectors in $R(T)$, and if k is any scalar, then there exist vectors $\mathbf{a}$ and $\mathbf{b}$ in V for which

$$T(\mathbf{a}) = \mathbf{w}_1+\mathbf{w}_2 \quad \text{and} \quad T(\mathbf{b}) = k\mathbf{w}_1 \tag{4}$$

But the fact that $\mathbf{w}_1$ and $\mathbf{w}_2$ are in $R(T)$ tells us there exist vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in V such that

$$T(\mathbf{v}_1) = \mathbf{w}_1 \quad \text{and} \quad T(\mathbf{v}_2) = \mathbf{w}_2$$

The following computations complete the proof by showing that the vectors $\mathbf{a} = \mathbf{v}_1+\mathbf{v}_2$ and $\mathbf{b} = k\mathbf{v}_1$ satisfy the equations in (4):

$$T(\mathbf{a}) = T(\mathbf{v}_1+\mathbf{v}_2) = T(\mathbf{v}_1)+T(\mathbf{v}_2) = \mathbf{w}_1+\mathbf{w}_2$$

$$T(\mathbf{b}) = T(k\mathbf{v}_1) = kT(\mathbf{v}_1) = k\mathbf{w}_1 \ \blacksquare$$

EXAMPLE 19 | Application to Differential Equations

CALCULUS REQUIRED

Differential equations of the form

$$y'' + \omega^2 y = 0 \qquad (\omega \text{ a positive constant}) \tag{5}$$

arise in the study of vibrations. The set of all solutions of this equation on the interval $(-\infty, \infty)$ is the kernel of the linear transformation $D : C^2(-\infty, \infty) \to C(-\infty, \infty)$, given by

$$D(y) = y'' + \omega^2 y$$

It is proved in standard textbooks on differential equations that the kernel is a two-dimensional subspace of $C^2(-\infty, \infty)$, so that if we can find two linearly independent solutions of (5), then all other solutions can be expressed as linear combinations of those two. We leave it for you to confirm by differentiating that

$$y_1 = \cos \omega x \quad \text{and} \quad y_2 = \sin \omega x$$

are solutions of (5). These functions are linearly independent since neither is a scalar multiple of the other, and thus

$$y = c_1 \cos \omega x + c_2 \sin \omega x \tag{6}$$

is a "general solution" of (5) in the sense that every choice of c_1 and c_2 produces a solution, and every solution is of this form.

Rank and Nullity of Linear Transformations

In Definition 1 of Section 4.9 we defined the notions of *rank* and *nullity* for an $m \times n$ matrix, and in Theorem 4.9.2, which we called the *Dimension Theorem for Matrices*, we proved that the sum of the rank and nullity is n. We will show next that this result is a special case of a more general result about linear transformations. We start with the following definition.

Definition 3

Let $T : V \to W$ be a linear transformation. In the case that the range of T is finite-dimensional its dimension is called the ***rank of T***; and if the kernel of T is finite-dimensional, then its dimension is called the ***nullity of T***. These dimensions are denoted, respectively, by

$$\operatorname{rank}(T) \quad \text{and} \quad \operatorname{nullity}(T)$$

The following theorem, whose proof is optional, generalizes Theorem 4.9.2.

Theorem 8.1.4

Dimension Theorem for Linear Transformations

If $T : V \to W$ is a linear transformation from a finite-dimensional vector space V to a vector space W, then the range of T is finite-dimensional, and

$$\operatorname{rank}(T) + \operatorname{nullity}(T) = \dim(V) \tag{7}$$

In the special case where A is an $m \times n$ matrix and $T_A : R^n \to R^m$ is multiplication by A, the kernel of T_A is the null space of A, and the range of T_A is the column space of A. Thus, it follows from Theorem 8.1.4 that

$$\operatorname{rank}(T_A) + \operatorname{nullity}(T_A) = n$$

OPTIONAL: Proof of Theorem 8.1.4 Assume that V is n-dimensional. We must show that

$$\dim(R(T)) + \dim(\ker(T)) = n$$

We will give the proof for the case where $1 \le \dim(\ker(T)) < n$. The cases where $\dim(\ker(T)) = 0$ and $\dim(\ker(T)) = n$ are left as exercises. Assume $\dim(\ker(T)) = r$, and let $\mathbf{v}_1, \dots, \mathbf{v}_r$ be a basis for the kernel. Since $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is linearly independent, Theorem 4.6.5(b) states that there are $n - r$ vectors, $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$, such that the extended set $\{\mathbf{v}_1, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ is a basis for V. To complete the proof, we will show that the $n - r$ vectors in the set $S = \{T(\mathbf{v}_{r+1}), \dots, T(\mathbf{v}_n)\}$ form a basis for the range of T. It will then follow that

$$\dim(R(T)) + \dim(\ker(T)) = (n - r) + r = n$$

First we show that S spans the range of T. If $\mathbf{b}$ is any vector in the range of T, then $\mathbf{b} = T(\mathbf{v})$ for some vector $\mathbf{v}$ in V. Since $\{\mathbf{v}_1, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ is a basis for V, the vector $\mathbf{v}$ can be written in the form

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r + c_{r+1}\mathbf{v}_{r+1} + \cdots + c_n\mathbf{v}_n$$

Since $\mathbf{v}_1, \dots, \mathbf{v}_r$ lie in the kernel of T, we have $T(\mathbf{v}_1) = \cdots = T(\mathbf{v}_r) = \mathbf{0}$, so

$$\mathbf{b} = T(\mathbf{v}) = c_{r+1}T(\mathbf{v}_{r+1}) + \cdots + c_nT(\mathbf{v}_n)$$

Thus S spans the range of T.

Finally, we show that S is a linearly independent set and consequently forms a basis for the range of T. Suppose that some linear combination of the vectors in S is zero; that is,

$$k_{r+1}T(\mathbf{v}_{r+1}) + \cdots + k_nT(\mathbf{v}_n) = \mathbf{0} \tag{8}$$

We must show that $k_{r+1} = \cdots = k_n = 0$. Since T is linear, (8) can be rewritten as

$$T(k_{r+1}\mathbf{v}_{r+1} + \cdots + k_n\mathbf{v}_n) = \mathbf{0}$$

which says that $k_{r+1}\mathbf{v}_{r+1} + \cdots + k_n\mathbf{v}_n$ is in the kernel of T. This vector can therefore be written as a linear combination of the basis vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$, say

$$k_{r+1}\mathbf{v}_{r+1} + \cdots + k_n\mathbf{v}_n = k_1\mathbf{v}_1 + \cdots + k_r\mathbf{v}_r$$

Thus,

$$k_1\mathbf{v}_1 + \cdots + k_r\mathbf{v}_r - k_{r+1}\mathbf{v}_{r+1} - \cdots - k_n\mathbf{v}_n = \mathbf{0}$$

Since $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly independent, all of the k's are zero; in particular, $k_{r+1} = \cdots = k_n = 0$, which completes the proof. ■

Exercise Set 8.1

In Exercises 1–2, suppose that T is a mapping whose domain is the vector space M_{22}. In each part, determine whether T is a linear transformation, and if so, find its kernel.

1. a. $T(A) = A^2$ b. $T(A) = \text{tr}(A)$
   c. $T(A) = A + A^T$

2. a. $T(A) = (A)_{11}$ b. $T(A) = 0_{2\times 2}$
   c. $T(A) = cA$

In Exercises 3–9, determine whether the mapping T is a linear transformation, and if so, find its kernel.

3. $T: R^3 \to R$, where $T(\mathbf{u}) = \|\mathbf{u}\|$.

4. $T: R^3 \to R^3$, where $\mathbf{v}_0$ is a fixed vector in R^3 and
$$T(\mathbf{u}) = \mathbf{u} \times \mathbf{v}_0$$

5. $T: M_{22} \to M_{23}$, where B is a fixed 2×3 matrix and
$$T(A) = AB$$

6. $T: M_{22} \to R$, where
   a. $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = 3a - 4b + c - d$
   b. $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a^2 + b^2$

7. $T: P_2 \to P_2$, where
   a. $T(a_0 + a_1x + a_2x^2) = a_0 + a_1(x + 1) + a_2(x + 1)^2$
   b. $T(a_0 + a_1x + a_2x^2) = (a_0 + 1) + (a_1 + 1)x + (a_2 + 1)x^2$

8. $T: F(-\infty, \infty) \to F(-\infty, \infty)$, where
   a. $T(f(x)) = 1 + f(x)$ b. $T(f(x)) = f(x + 1)$

9. $T: R^\infty \to R^\infty$, where

$$T(a_0, a_1, a_2, \ldots, a_n, \ldots) = (0, a_0, a_1, a_2, \ldots, a_n, \ldots)$$

10. Let $T: P_2 \to P_3$ be the linear transformation defined by $T(p(x)) = xp(x)$. Which of the following are in $\ker(T)$?

a. x^2 **b.** 0 **c.** $1 + x$ **d.** $-x$

11. Let $T: P_2 \to P_3$ be the linear transformation in Exercise 10. Which of the following are in $R(T)$?

a. $x + x^2$ **b.** $1 + x$ **c.** $3 - x^2$ **d.** $-x$

12. Let V be any vector space, and let $T: V \to V$ be defined by $T(\mathbf{v}) = 3\mathbf{v}$.

a. What is the kernel of T?

b. What is the range of T?

13. In each part, use the given information to find the nullity of the linear transformation T.

a. $T: R^5 \to P_5$ has rank 3.

b. $T: P_4 \to P_3$ has rank 1.

c. The range of $T: M_{mn} \to R^3$ is R^3.

d. $T: M_{22} \to M_{22}$ has rank 3.

14. In each part, use the given information to find the rank of the linear transformation T.

a. $T: R^7 \to M_{32}$ has nullity 2.

b. $T: P_3 \to R$ has nullity 1.

c. The null space of $T: P_5 \to P_5$ is P_5.

d. $T: P_n \to M_{mn}$ has nullity 3.

15. Let $T: M_{22} \to M_{22}$ be the dilation operator with factor $k = 3$.

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

b. Find the rank and nullity of T.

16. Let $T: P_2 \to P_2$ be the contraction operator with factor $k = 1/4$.

a. Find $T(1 + 4x + 8x^2)$.

b. Find the rank and nullity of T.

17. Let $T: P_2 \to R^3$ be the evaluation transformation at the sequence of points $-1, 0, 1$. Find

a. $T(x^2)$ **b.** $\ker(T)$ **c.** $R(T)$

18. Let V be the subspace of $C[0, 2\pi]$ spanned by the vectors 1, $\sin x$, and $\cos x$, and let $T: V \to R^3$ be the evaluation transformation at the sequence of points $0, \pi, 2\pi$. Find

a. $T(1 + \sin x + \cos x)$ **b.** $\ker(T)$

c. $R(T)$

19. Consider the basis $S = \{\mathbf{v}_1, \mathbf{v}_2\}$ for R^2, where $\mathbf{v}_1 = (1, 1)$ and $\mathbf{v}_2 = (1, 0)$, and let $T: R^2 \to R^2$ be the linear operator for which

$$T(\mathbf{v}_1) = (1, -2) \quad \text{and} \quad T(\mathbf{v}_2) = (-4, 1)$$

Find a formula for $T(x_1, x_2)$, and use that formula to find $T(5, -3)$.

20. Consider the basis $S = \{\mathbf{v}_1, \mathbf{v}_2\}$ for R^2, where $\mathbf{v}_1 = (-2, 1)$ and $\mathbf{v}_2 = (1, 3)$, and let $T: R^2 \to R^3$ be the linear transformation such that

$$T(\mathbf{v}_1) = (-1, 2, 0) \quad \text{and} \quad T(\mathbf{v}_2) = (0, -3, 5)$$

Find a formula for $T(x_1, x_2)$, and use that formula to find $T(2, -3)$.

21. Consider the basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for R^3, where $\mathbf{v}_1 = (1, 1, 1)$, $\mathbf{v}_2 = (1, 1, 0)$, and $\mathbf{v}_3 = (1, 0, 0)$, and let $T: R^3 \to R^3$ be the linear operator for which

$$T(\mathbf{v}_1) = (2, -1, 4), \quad T(\mathbf{v}_2) = (3, 0, 1),$$
$$T(\mathbf{v}_3) = (-1, 5, 1)$$

Find a formula for $T(x_1, x_2, x_3)$, and use that formula to find $T(2, 4, -1)$.

22. Consider the basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for R^3, where $\mathbf{v}_1 = (1, 2, 1)$, $\mathbf{v}_2 = (2, 9, 0)$, and $\mathbf{v}_3 = (3, 3, 4)$, and let $T: R^3 \to R^2$ be the linear transformation for which

$$T(\mathbf{v}_1) = (1, 0), \quad T(\mathbf{v}_2) = (-1, 1), \quad T(\mathbf{v}_3) = (0, 1)$$

Find a formula for $T(x_1, x_2, x_3)$, and use that formula to find $T(7, 13, 7)$.

In Exercises 23–24, let T be multiplication by the matrix A. Find

a. a basis for the range of T.

b. a basis for the kernel of T.

c. the rank and nullity of T.

d. the rank and nullity of A.

23. $A = \begin{bmatrix} 1 & -1 & 3 \\ 5 & 6 & -4 \\ 7 & 4 & 2 \end{bmatrix}$ **24.** $A = \begin{bmatrix} 2 & 0 & -1 \\ 4 & 0 & -2 \\ 20 & 0 & 0 \end{bmatrix}$

In Exercises 25–26, let $T_A: R^4 \to R^3$ be multiplication by A. Find a basis for the kernel of T_A, and then find a basis for the range of T_A that consists of column vectors of A.

25. $A = \begin{bmatrix} 1 & 2 & -1 & -2 \\ -3 & 1 & 3 & 4 \\ -3 & 8 & 4 & 2 \end{bmatrix}$

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

27. Let $T: P_3 \to P_2$ be the mapping defined by

$$T(a_0 + a_1x + a_2x^2 + a_3x^3) = 5a_0 + a_3x^2$$

a. Show that T is linear.

b. Find a basis for the kernel of T.

c. Find a basis for the range of T.

28. Let $T: P_2 \to P_2$ be the mapping defined by

$$T(a_0 + a_1x + a_2x^2) = 3a_0 + a_1x + (a_0 + a_1)x^2$$

a. Show that T is linear.

b. Find a basis for the kernel of T.

c. Find a basis for the range of T.

29. **a.** **(*Calculus required*)** Let $D: P_3 \to P_2$ be the differentiation transformation $D(\mathbf{p}) = p'(x)$. What is the kernel of D?

b. **(*Calculus required*)** Let $J: P_1 \to R$ be the integration transformation $J(\mathbf{p}) = \int_{-1}^{1} p(x)\,dx$. What is the kernel of J?

30. **(*Calculus required*)** Let $V = C[a, b]$ be the vector space of continuous functions on $[a, b]$, and let $T: V \to V$ be the transformation defined by

$$T(\mathbf{f}) = 5f(x) + 3\int_a^x f(t)\,dt$$

Is T a linear operator?

31. **(*Calculus required*)** Let V be the vector space of real-valued functions with continuous derivatives of all orders on the interval $(-\infty, \infty)$, and let $W = F(-\infty, \infty)$ be the vector space of real-valued functions defined on $(-\infty, \infty)$.

a. Find a linear transformation $T: V \to W$ whose kernel is P_3.

b. Find a linear transformation $T: V \to W$ whose kernel is P_n.

32. For a positive integer $n > 1$, let $T: M_{nn} \to R$ be the linear transformation defined by $T(A) = \operatorname{tr}(A)$, where A is an $n \times n$ matrix with real entries. Determine the dimension of $\ker(T)$.

33. **a.** Let $T: V \to R^3$ be a linear transformation from a vector space V to R^3. Geometrically, what are the possibilities for the range of T?

b. Let $T: R^3 \to W$ be a linear transformation from R^3 to a vector space W. Geometrically, what are the possibilities for the kernel of T?

34. In each part, determine whether the mapping $T: P_n \to P_n$ is linear.

a. $T(p(x)) = p(x + 1)$

b. $T(p(x)) = p(x) + 1$

35. Let $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ be vectors in a vector space V, and let $T: V \to R^3$ be a linear transformation for which

$$T(\mathbf{v}_1) = (1, -1, 2), \quad T(\mathbf{v}_2) = (0, 3, 2),$$
$$T(\mathbf{v}_3) = (-3, 1, 2)$$

Find $T(2\mathbf{v}_1 - 3\mathbf{v}_2 + 4\mathbf{v}_3)$.

Working with Proofs

36. Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for a vector space V, and let $T: V \to W$ be a linear transformation. Prove that if

$$T(\mathbf{v}_1) = T(\mathbf{v}_2) = \cdots = T(\mathbf{v}_n) = \mathbf{0}$$

then T is the zero transformation.

37. Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for a vector space V, and let $T: V \to V$ be a linear operator. Prove that if

$$T(\mathbf{v}_1) = \mathbf{v}_1, \quad T(\mathbf{v}_2) = \mathbf{v}_2, \ldots, \quad T(\mathbf{v}_n) = \mathbf{v}_n$$

then T is the identity transformation on V.

38. Prove: If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for a vector space V and $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ are vectors in a vector space W, not necessarily distinct, then there exists a linear transformation T that maps V into W such that

$$T(\mathbf{v}_1) = \mathbf{w}_1, \quad T(\mathbf{v}_2) = \mathbf{w}_2, \ldots, \quad T(\mathbf{v}_n) = \mathbf{w}_n$$

39. Let $q_0(x)$ be a fixed polynomial of degree m, and define a function T with domain P_n by the formula $T(p(x)) = p(q_0(x))$. Prove that T is a linear transformation.

True-False Exercises

TF. In parts **(a)–(i)** determine whether the statement is true or false, and justify your answer.

a. If $T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2)$ for all vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in V and all scalars c_1 and c_2, then T is a linear transformation.

b. If $\mathbf{v}$ is a nonzero vector in V, then there is exactly one linear transformation $T: V \to W$ such that

$$T(-\mathbf{v}) = -T(\mathbf{v})$$

c. There is exactly one linear transformation $T: V \to W$ for which $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u} - \mathbf{v})$ for all vectors $\mathbf{u}$ and $\mathbf{v}$ in V.

d. If $\mathbf{v}_0$ is a nonzero vector in V, then $T(\mathbf{v}) = \mathbf{v}_0 + \mathbf{v}$ defines a linear operator on V.

e. The kernel of a linear transformation is a vector space.

f. The range of a linear transformation is a vector space.

g. If $T: P_6 \to M_{22}$ is a linear transformation, then the nullity of T is 3.

h. The function $T: M_{22} \to R$ defined by $T(A) = \det A$ is a linear transformation.

i. The linear transformation $T: M_{22} \to M_{22}$ defined by

$$T(A) = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} A$$

has rank 1.

8.2 Compositions and Inverse Transformations

In Section 1.9 we discussed compositions and inverses of matrix transformations. In this section we will extend some of those ideas to general linear transformations.

One-to-One and Onto

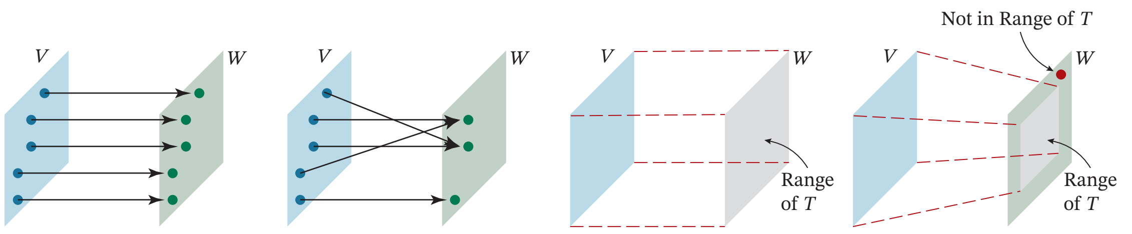
To set the groundwork for our discussion in this section we will need the following definitions that are illustrated in **Figure 8.2.1**.

Definition 1

If $T: V \to W$ is a linear transformation from a vector space V to a vector space W, then T is said to be ***one-to-one*** if T maps distinct vectors in V into distinct vectors in W.

Definition 2

If $T: V \to W$ is a linear transformation from a vector space V to a vector space W, then T is said to be ***onto*** (or ***onto*** $\boldsymbol{W}$) if every vector in W is the image of at least one vector in V.



One-to-one. Distinct vectors in V have distinct images in W.

Not one-to-one. There exist distinct vectors in V with the same image.

Onto W. Every vector in W is the image of some vector in V.

Not onto W. Not every vector in W is the image of some vector in V.

FIGURE 8.2.1

The idea of a one-to-one linear transformation can be expressed in other ways as well:

1. $T: V \to W$ is one-to-one if and only if for each vector $\mathbf{w}$ in the range of T, there is exactly one vector $\mathbf{v}$ in V such that $T(\mathbf{v}) = \mathbf{w}$.
2. $T: V \to W$ is one-to-one if and only if $T(\mathbf{u}) = T(\mathbf{v})$ implies that $\mathbf{u} = \mathbf{v}$.

Recall from Definition 2 of Section 8.1 that the *kernel* of a linear transformation consists of all vectors that the transformation maps into $\mathbf{0}$. The following theorem links that definition with the concept of a one-to-one linear transformation.

Theorem 8.2.1

If $T : V \to W$ is a linear transformation, then the following two statements are equivalent.

(*a*) T is one-to-one.

(*b*) $\ker(T) = \{\mathbf{0}\}$.

Proof (a) $\Rightarrow$ ***(b)*** Since T is linear, we know that $T(\mathbf{0}) = \mathbf{0}$ by Theorem 8.1.1(*a*). Since T is one-to-one, there can be no other vectors in V that map into $\mathbf{0}$, so $\ker(T) = \{\mathbf{0}\}$.

(b) $\Rightarrow$ ***(a)*** Assume that $\ker(T) = \{\mathbf{0}\}$. If $\mathbf{u}$ and $\mathbf{v}$ are distinct vectors in V, then $\mathbf{u} - \mathbf{v} \neq \mathbf{0}$. This implies that $T(\mathbf{u} - \mathbf{v}) \neq \mathbf{0}$, for otherwise $\ker(T)$ would contain a nonzero vector. Since T is linear, it follows that

$$T(\mathbf{u}) - T(\mathbf{v}) = T(\mathbf{u} - \mathbf{v}) \neq \mathbf{0}$$

so T maps distinct vectors in V into distinct vectors in W and hence is one-to-one. ■

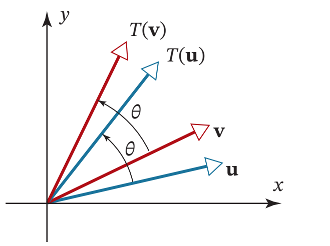


FIGURE 8.2.2 Distinct vectors $\mathbf{u}$ and $\mathbf{v}$ are rotated into distinct vectors $T(\mathbf{u})$ and $T(\mathbf{v})$.

EXAMPLE 1 | Rotation Operators on R^2 Are One-to-One and Onto

The linear operator $T : R^2 \to R^2$ that rotates each vector in the plane about the origin through an angle θ is one-to-one because it maps distinct vectors into distinct vectors (**Figure 8.2.2**). It is also onto because every vector in R^2 is the image under this rotation of another vector in R^2 (which vector?).

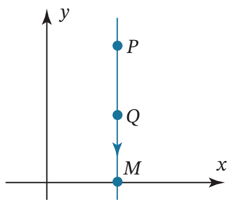


FIGURE 8.2.3 The distinct points P and Q are mapped into the same point M.

EXAMPLE 2 | Orthogonal Projections in R^2 Are Not One-to-One

The linear operator $T : R^2 \to R^2$ that maps points orthogonally on to the x-axis in R^2 maps distinct points on a vertical line to the same point on the x-axis and hence is not one-to-one (**Figure 8.2.3**). It is also not onto R^2 because points off the x-axis are not images of any point in R^2 under such a projection. Similarly, orthogonal projections onto the y-axis are neither one-to-one nor onto.

EXAMPLE 3 | Two Transformations That Are One-to-One and Onto

The linear transformations $T_1 : P_3 \to R^4$ and $T_2 : M_{22} \to R^4$ defined by

$$T_1(a + bx + cx^2 + dx^3) = (a, b, c, d)$$

$$T_2\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a, b, c, d)$$

are both onto R^4 because every vector in R^4 can be obtained by choosing a, b, c, and d appropriately. Both transformations are one-to-one because their kernels contain only the zero vector in their respective domains (verify).

EXAMPLE 4 | A One-to-One Linear Transformation That Is Not Onto

Let $T: P_n \to P_{n+1}$ be the linear transformation

$$T(\mathbf{p}) = T(p(x)) = xp(x)$$

discussed in Example 5 of Section 8.1. If

$$\mathbf{p} = p(x) = c_0 + c_1 x + \cdots + c_n x^n \quad \text{and} \quad \mathbf{q} = q(x) = d_0 + d_1 x + \cdots + d_n x^n$$

are distinct polynomials, then they differ in at least one coefficient, and hence

$$T(\mathbf{p}) = c_0 x + c_1 x^2 + \cdots + c_n x^{n+1} \quad \text{and} \quad T(\mathbf{q}) = d_0 x + d_1 x^2 + \cdots + d_n x^{n+1}$$

also differ in at least one coefficient. Thus, T is one-to-one, since it maps distinct polynomials into distinct polynomials. However, it is not onto because all images under T have a zero constant term, and hence there is no polynomial in P_n that maps into the constant polynomial 1.

EXAMPLE 5 | Shifting Operators

Let $V = R^\infty$ be the sequence space discussed in Example 3 of Section 4.1, and consider the linear "shifting operators" on V defined by

$$T_1(u_1, u_2, \ldots, u_n, \ldots) = (0, u_1, u_2, \ldots, u_n, \ldots)$$
$$T_2(u_1, u_2, \ldots, u_n, \ldots) = (u_2, u_3, \ldots, u_n, \ldots)$$

(*a*) Show that T_1 is one-to-one but not onto.

(*b*) Show that T_2 is onto but not one-to-one.

Solution (*a*) The operator T_1 is one-to-one because distinct sequences in R^∞ obviously have distinct images. This operator is not onto because no vector in R^∞ maps into the sequence $(1, 0, 0, \ldots, 0, \ldots)$, for example.

Solution (*b*) The operator T_2 is not one-to-one because, for example, the distinct vectors $(1, 0, 0, \ldots, 0, \ldots)$ and $(2, 0, 0, \ldots, 0, \ldots)$ both map into $(0, 0, 0, \ldots, 0, \ldots)$. This operator is onto because every possible sequence of real numbers can be obtained with an appropriate choice of the numbers $u_2, u_3, \ldots, u_n, \ldots$.

EXAMPLE 6 | Differentiation Is Not One-to-One

CALCULUS REQUIRED

Let

$$D: C^1(-\infty, \infty) \to F(-\infty, \infty)$$

be the differentiation transformation discussed in Example 11 of Section 8.1. This linear transformation is *not* one-to-one because it maps functions that differ by a constant into the same function. For example,

$$D(x^2) = D(x^2 + 1) = 2x$$

In the special case where V and W are finite-dimensional and have the same dimension, we can add a third statement to those in Theorem 8.2.1.

Theorem 8.2.2

If V and W are finite-dimensional vector spaces with the same dimension, and if $T: V \to W$ is a linear transformation, then the following statements are equivalent.

(a) T is one-to-one.

(b) $\ker(T) = \{\mathbf{0}\}$.

(c) T is onto [i.e., $R(T) = W$].

Why does Example 5 not violate Theorem 8.2.2?

Proof We already know that (a) and (b) are equivalent by Theorem 8.2.1, so it suffices to show that (b) and (c) are equivalent. We leave it for you to do this by assuming that $\dim(V) = n$ and applying Theorem 8.1.4. ■

The requirement in Theorem 8.2.2 that V and W have the same dimension is essential for the validity of the theorem. In the exercises we will ask you to prove the following facts for the case where they do not have the same dimension.

- If $\dim(W) < \dim(V)$, then T cannot be one-to-one.
- If $\dim(V) < \dim(W)$, then T cannot be onto.

Stated informally, if a linear transformation maps a "bigger" space to a "smaller" space, then some points in the "bigger" space must have the same image; and if a linear transformation maps a "smaller" space to a "bigger" space, then there must be points in the "bigger" space that are not images of any points in the "smaller" space.

In retrospect, had Theorem 8.2.2 been available prior to Example 3, it would have sufficed to show that the transformations were *either* one-to-one *or* onto since P_3 and M_{22} have the same dimension as R^4 (dimension 4).

Matrix Transformations Revisited

Let us return for the moment to matrix transformations and consider an example that illustrates the two results about dimension that followed Theorem 8.2.2.

EXAMPLE 7 | Matrix Transformations from R^n to R^m

If $T_A: R^n \to R^m$ is multiplication by an $m \times n$ matrix A, then it follows from the discussion immediately following the proof of Theorem 8.2.2 that T_A is not one-to-one if $m < n$ and not onto if $n < m$. In the case where $m = n$, whether or not T_A is one-to-one or onto depends on the rank of the matrix A. However, in the exercises we will ask you to show that if A is invertible, then T_A will be both one-to-one and onto.

The following theorem illustrates that it is the column vectors of a matrix A that determine whether the matrix transformation $T_A: R^n \to R^m$ is one-to-one or onto.

Theorem 8.2.3

If $T_A : R^n \to R^m$ is a matrix transformation, then

(*a*) T_A is one-to-one if and only if the columns of A are linearly independent.

(*b*) T_A is onto if and only if the columns of A span R^m.

Proof (a) It follows from Theorem 8.2.1 that T_A is one-to-one if and only if A has nullity 0, which is equivalent to saying that A has rank m (Theorem 4.9.2), which is equivalent to saying that the m column vectors of A are linearly independent.

Proof (b) To say that T_A is onto is equivalent to saying that the system $A\mathbf{x} = \mathbf{b}$ has a solution for every vector $\mathbf{b}$ in R^m. But this is so if and only if every vector $\mathbf{b}$ in R^m is in the column space of A (Theorem 4.8.1), which is so if and only if the columns of A span R^m. ■

We leave it as an exercise to show that parts (t), (u), and (v) below can be added to Equivalence Theorem 8.2.4 in the case where $T_A : R^n \to R^n$ is a linear operator.

Theorem 8.2.4

Equivalent Statements

If A is an $n \times n$ matrix in which there are no duplicate rows and no duplicate columns, then the following statements are equivalent.

(a) A is invertible.

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

(c) The reduced row echelon form of A is I_n.

(d) A is expressible as a product of elementary matrices.

(e) $A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.

(f) $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.

(g) $\det(A) \neq 0$.

(h) The column vectors of A are linearly independent.

(i) The row vectors of A are linearly independent.

(j) The column vectors of A span R^n.

(k) The row vectors of A span R^n.

(l) The column vectors of A form a basis for R^n.

(m) The row vectors of A form a basis for R^n.

(n) A has rank n.

(o) A has nullity 0.

(p) The orthogonal complement of the null space of A is R^n.

(q) The orthogonal complement of the row space of A is $\{\mathbf{0}\}$.

(r) $\lambda = 0$ is not an eigenvalue of A.

(s) A^TA is invertible.

(t) The kernel of T_A is $\{\mathbf{0}\}$.

(u) The range of T_A is R^n.

(v) T_A is one-to-one.

The key to solving a mathematical problem is often adopting the right point of view; and this is why, in linear algebra, we develop different ways of thinking about the same vector space. For example, if A is an $m \times n$ matrix, here are three ways of viewing the same subspace of R^n:

- **Matrix view:** the null space of A
- **System view:** the solution space of $A\mathbf{x} = \mathbf{0}$
- **Transformation view:** the kernel of T_A

and here are three ways of viewing the same subspace of R^m:

- **Matrix view:** the column space of A
- **System view:** all $\mathbf{b}$ in R^m for which $A\mathbf{x} = \mathbf{b}$ is consistent
- **Transformation view:** the range of T_A

Inverse Linear Transformations

In Section 1.9 we introduced the concept of an invertible matrix operator, and in this subsection we will extend that idea to general linear transformations. By way of review, recall that a matrix operator $T_A : R^n \to R^n$ is invertible if and only if the matrix A is invertible, in which case the inverse of that operator is $T_{A^{-1}} : R^n \to R^n$. In words, *the inverse of multiplication by A is multiplication by A^{-1}*.

EXAMPLE 8 | A One-to-One Matrix Transformation

Let $T : R^3 \to R^3$ be the linear operator defined by the formula

$$T(x_1, x_2, x_3) = (3x_1 + x_2, -2x_1 - 4x_2 + 3x_3, 5x_1 + 4x_2 - 2x_3)$$

Determine whether T is one-to-one; if so, find $T^{-1}(x_1, x_2, x_3)$.

Solution The stated formula defines a matrix transformation whose standard matrix by Formula (15) of Section 1.8 is

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

(verify). This matrix is invertible and its inverse is

$$A^{-1} = \begin{bmatrix} 4 & -2 & -3 \\ -11 & 6 & 9 \\ -12 & 7 & 10 \end{bmatrix}$$

Thus, the transformation T is invertible and

$$T^{-1}\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = A^{-1}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 & -2 & -3 \\ -11 & 6 & 9 \\ -12 & 7 & 10 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4x_1 - 2x_2 - 3x_3 \\ -11x_1 + 6x_2 + 9x_3 \\ -12x_1 + 7x_2 + 10x_3 \end{bmatrix}$$

Expressing this result in comma delimited notation yields

$$T^{-1}(x_1, x_2, x_3) = (4x_1 - 2x_2 - 3x_3, -11x_1 + 6x_2 + 9x_3, -12x_1 + 7x_2 + 10x_3)$$

Now let us turn our attention to the invertibility of general linear transformations. If $T: V \to W$ is a one-to-one linear transformation with range $R(T)$, and if $\mathbf{w}$ is any vector in $R(T)$, then the fact that T is one-to-one means that there is *exactly one* vector $\mathbf{v}$ in V for which $T(\mathbf{v}) = \mathbf{w}$. This fact allows us to define a new function, called the ***inverse of T*** (and denoted by T^{-1}), that is defined on the range of T and that maps $\mathbf{w}$ back into $\mathbf{v}$ (**Figure 8.2.4**).

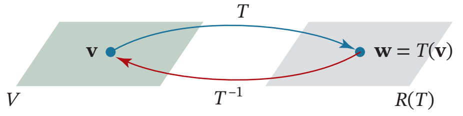


FIGURE 8.2.4 The inverse of T maps $T(\mathbf{v})$ back into $\mathbf{v}$.

In the exercises we will ask you to prove that $T^{-1}: R(T) \to V$ is a linear transformation. Moreover, it follows from the definition of T^{-1} that

$$\boxed{T^{-1}(T(\mathbf{v})) = T^{-1}(\mathbf{w}) = \mathbf{v}} \tag{1}$$

$$\boxed{T(T^{-1}(\mathbf{w})) = T(\mathbf{v}) = \mathbf{w}} \tag{2}$$

so that T and T^{-1}, when applied in succession in either order, cancel the effect of each other.

EXAMPLE 9 | An Inverse Transformation

We showed in Example 4 of this section that the linear transformation $T: P_n \to P_{n+1}$ given by

$$T(\mathbf{p}) = T(p(x)) = xp(x)$$

is one-to-one but not onto. The fact that it is not onto can be seen explicitly from the formula

$$T(c_0 + c_1x + \cdots + c_nx^n) = c_0x + c_1x^2 + \cdots + c_nx^{n+1} \tag{3}$$

The fact that T is not onto does not preclude the existence of an inverse, since the inverse is defined on the range of T. It is evident from (3) the range in this case consists of all polynomials of degree $n + 1$ or less that have a zero constant term and that the inverse is given by the formula

$$T^{-1}(c_0x + c_1x^2 + \cdots + c_nx^{n+1}) = c_0 + c_1x + \cdots + c_nx^n$$

For example, in the case where $n \geq 3$,

$$T^{-1}(2x - x^2 + 5x^3 + 3x^4) = 2 - x + 5x^2 + 3x^3$$

Composition of Linear Transformations

The following definition extends Formula (1) of Section 1.9 to general linear transformations.

Definition 3

If $T_1: U \to V$ and $T_2: V \to W$ are linear transformations, then the ***composition*** of T_2 with T_1, denoted by $T_2 \circ T_1$ (and which is read "T_2 circle T_1"), is the function defined by the formula

$$\boxed{(T_2 \circ T_1)(\mathbf{u}) = T_2(T_1(\mathbf{u}))} \tag{4}$$

where $\mathbf{u}$ is a vector in U.

Note that the word "with" establishes the order of the operations in a composition. The composition of T_2 *with* T_1 is

$$(T_2 \circ T_1)(\mathbf{u}) = T_2(T_1(\mathbf{u}))$$

whereas the composition of T_1 *with* T_2 is

$$(T_1 \circ T_2)(\mathbf{u}) = T_1(T_2(\mathbf{u}))$$

It is not true, in general, that $T_1 \circ T_2 = T_2 \circ T_1$.

Remark Observe that this definition requires that the domain of T_2 (which is V) contain the range of T_1. This is essential for the formula $T_2(T_1(\mathbf{u}))$ to make sense (**Figure 8.2.5**).

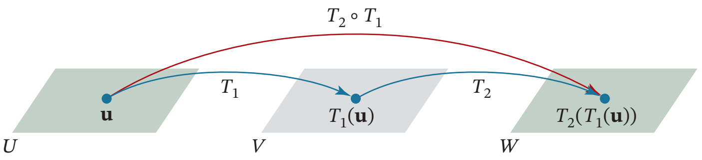


FIGURE 8.2.5 The composition of T_2 with T_1.

Our next theorem shows that the composition of two linear transformations is itself a linear transformation.

Theorem 8.2.5

If $T_1 : U \to V$ and $T_2 : V \to W$ are linear transformations, then $(T_2 \circ T_1) : U \to W$ is also a linear transformation.

Proof If $\mathbf{u}$ and $\mathbf{v}$ are vectors in U and c is a scalar, then it follows from (4) and the linearity of T_1 and T_2 that

$$\begin{aligned}(T_2 \circ T_1)(\mathbf{u}+\mathbf{v}) &= T_2(T_1(\mathbf{u}+\mathbf{v})) = T_2(T_1(\mathbf{u}) + T_1(\mathbf{v})) \\ &= T_2(T_1(\mathbf{u})) + T_2(T_1(\mathbf{v})) \\ &= (T_2 \circ T_1)(\mathbf{u}) + (T_2 \circ T_1)(\mathbf{v})\end{aligned}$$

and

$$\begin{aligned}(T_2 \circ T_1)(c\mathbf{u}) &= T_2(T_1(c\mathbf{u})) = T_2(cT_1(\mathbf{u})) \\ &= cT_2(T_1(\mathbf{u})) = c(T_2 \circ T_1)(\mathbf{u})\end{aligned}$$

Thus, $T_2 \circ T_1$ satisfies the two requirements of a linear transformation. ■

EXAMPLE 10 | Composition of Linear Transformations

Let $T_1 : P_1 \to P_2$ and $T_2 : P_2 \to P_2$ be the linear transformations given by the formulas

$$T_1(p(x)) = xp(x) \quad \text{and} \quad T_2(p(x)) = p(2x+4)$$

Then the composition $(T_2 \circ T_1) : P_1 \to P_2$ is given by the formula

$$(T_2 \circ T_1)(p(x)) = T_2(T_1(p(x))) = T_2(xp(x)) = (2x+4)p(2x+4)$$

In particular, if $p(x) = c_0 + c_1 x$, then

$$\begin{aligned}(T_2 \circ T_1)(p(x)) = (T_2 \circ T_1)(c_0 + c_1 x) &= (2x+4)(c_0 + c_1(2x+4)) \\ &= c_0(2x+4) + c_1(2x+4)^2\end{aligned}$$

EXAMPLE 11 | Composition with the Identity Operator

If $T : V \to V$ is any linear operator, and if $I : V \to V$ is the identity operator (Example 3 of Section 8.1), then for all vectors $\mathbf{v}$ in V, we have

$$(T \circ I)(\mathbf{v}) = T(I(\mathbf{v})) = T(\mathbf{v})$$
$$(I \circ T)(\mathbf{v}) = I(T(\mathbf{v})) = T(\mathbf{v})$$

It follows that $T \circ I$ and $I \circ T$ are the same as T; that is,

$$\boxed{T \circ I = T} \quad \text{and} \quad \boxed{I \circ T = T} \tag{5}$$

As illustrated in **Figure 8.2.6**, compositions can be defined for more than two linear transformations. For example, if

$$T_1 : U \to V, \quad T_2 : V \to W, \quad \text{and} \quad T_3 : W \to Y$$

are linear transformations, then the composition $T_3 \circ T_2 \circ T_1$ is defined by

$$\boxed{(T_3 \circ T_2 \circ T_1)(\mathbf{u}) = T_3(T_2(T_1(\mathbf{u})))} \tag{6}$$

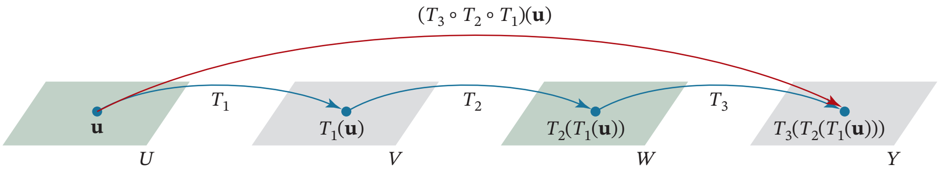


FIGURE 8.2.6 The composition of three linear transformations.

Composition of One-to-One Linear Transformations

Our next theorem shows that the composition of one-to-one linear transformations is one-to-one and that the inverse of a composition is the composition of the inverses in the reverse order.

Theorem 8.2.6

If $T_1 : U \to V$ and $T_2 : V \to W$ are one-to-one linear transformations, then:

(*a*) $T_2 \circ T_1$ is one-to-one.

(*b*) $(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}$.

Note the order of the subscripts on the two sides of the formula in part (*b*) of Theorem 8.2.5.

Proof (a) We want to show that $T_2 \circ T_1$ maps distinct vectors in U into distinct vectors in W. But if $\mathbf{u}$ and $\mathbf{v}$ are distinct vectors in U, then $T_1(\mathbf{u})$ and $T_1(\mathbf{v})$ are distinct vectors in V since T_1 is one-to-one. This and the fact that T_2 is one-to-one imply that

$$T_2(T_1(\mathbf{u})) \quad \text{and} \quad T_2(T_1(\mathbf{v}))$$

are also distinct vectors. But these expressions can also be written as

$$(T_2 \circ T_1)(\mathbf{u}) \quad \text{and} \quad (T_2 \circ T_1)(\mathbf{v})$$

so $T_2 \circ T_1$ maps $\mathbf{u}$ and $\mathbf{v}$ into distinct vectors in W.

Proof (b) We want to show that

$$(T_2 \circ T_1)^{-1}(\mathbf{w}) = (T_1^{-1} \circ T_2^{-1})(\mathbf{w})$$

for every vector $\mathbf{w}$ in the range of $T_2 \circ T_1$. For this purpose, let

$$\mathbf{u} = (T_2 \circ T_1)^{-1}(\mathbf{w}) \tag{7}$$

so our goal is to show that

$$\mathbf{u} = (T_1^{-1} \circ T_2^{-1})(\mathbf{w})$$

But it follows from (7) that

$$(T_2 \circ T_1)(\mathbf{u}) = \mathbf{w}$$

or, equivalently,

$$T_2(T_1(\mathbf{u})) = \mathbf{w}$$

Now, taking T_2^{-1} of each side of this equation, then taking T_1^{-1} of each side of the result, and then using (1) yields (verify)

$$\mathbf{u} = T_1^{-1}(T_2^{-1}(\mathbf{w}))$$

or, equivalently,

$$\mathbf{u} = (T_1^{-1} \circ T_2^{-1})(\mathbf{w}) \blacksquare$$

In words, part (*b*) of Theorem 8.2.5 states that *the inverse of a composition is the composition of the inverses in the reverse order*. This result can be extended to compositions of three or more linear transformations; for example,

$$\boxed{(T_3 \circ T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1} \circ T_3^{-1}} \tag{8}$$

Part (*b*) of Theorem 8.2.5 and Formula (8) apply to general linear transformations. In the special case where they are matrix transformations they can be written as

$$(T_B \circ T_A)^{-1} = T_A^{-1} \circ T_B^{-1} \quad \text{and} \quad (T_C \circ T_B \circ T_A)^{-1} = T_A^{-1} \circ T_B^{-1} \circ T_C^{-1}$$

or equivalently as

$$\boxed{(T_{BA})^{-1} = T_{A^{-1}B^{-1}} \quad \text{and} \quad (T_{CBA})^{-1} = T_{A^{-1}B^{-1}C^{-1}}} \tag{9}$$

Exercise Set 8.2

In Exercises 1–2, determine whether the stated matrix operator is one-to-one.

1. **a.** The orthogonal projection onto the x-axis in R^2.

b. The reflection about the y-axis in R^2.

c. The reflection about the line $y = x$ in R^2.

2. **a.** A rotation about the z-axis in R^3.

b. A reflection about the xy-plane in R^3.

c. An orthogonal projection onto the xz-plane in R^3.

In Exercises 3–4, determine whether the linear transformation is one-to-one by finding its kernel and then applying Theorem 8.2.1.

3. **a.** $T: R^2 \to R^2$, where $T(x, y) = (y, x)$

b. $T: R^2 \to R^3$, where $T(x, y) = (x, y, x + y)$

c. $T: R^3 \to R^2$, where $T(x, y, z) = (x + y + z, x - y - z)$

4. **a.** $T: R^2 \to R^3$, where $T(x, y) = (x - y, y - x, 2x - 2y)$

b. $T: R^2 \to R^2$, where $T(x, y) = (0, 2x + 3y)$

c. $T: R^2 \to R^2$, where $T(x, y) = (x + y, x - y)$

In Exercises 5–6, determine whether multiplication by A is one-to-one by computing the nullity of A and then applying Theorem 8.2.1.

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

b. $A = \begin{bmatrix} 1 & 3 & 1 & 7 \\ 2 & 7 & 2 & 4 \\ -1 & -3 & 0 & 0 \end{bmatrix}$

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

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

7. Use the given information to determine whether the linear transformation is one-to-one.

a. $T : V \to W$; nullity$(T) = 0$

b. $T : V \to W$; rank$(T) = \dim(V)$

c. $T : V \to W$; $\dim(W) < \dim(V)$

8. Use the given information to determine whether the linear operator is one-to-one, onto, both, or neither.

a. $T : V \to V$; nullity$(T) = 0$

b. $T : V \to V$; rank$(T) < \dim(V)$

c. $T : V \to V$; $R(T) = V$

9. Show that the linear transformation $T : P_2 \to R^2$ defined by $T(p(x)) = (p(-1), p(1))$ is not one-to-one by finding a nonzero polynomial that maps into $\mathbf{0} = (0, 0)$. Do you think that this transformation is onto?

10. Show that the linear transformation $T : P_2 \to P_2$ defined by $T(p(x)) = p(x+1)$ is one-to-one. Do you think that this transformation is onto?

11. Let $\mathbf{a}$ be a fixed vector in R^3. Does the formula $T(\mathbf{v}) = \mathbf{a} \times \mathbf{v}$ define a one-to-one linear operator on R^3? Explain your reasoning.

12. Let E be a fixed 2×2 elementary matrix. Does the formula $T(A) = EA$ define a one-to-one linear operator on M_{22}? Explain your reasoning.

In Exercises 13–14, use Theorem 8.2.3 to determine whether multiplication by A is one-to-one, onto, both, or neither. Justify your answer.

13. a. $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 5 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 1 & 3 & 1 & 1 \\ 1 & 3 & 1 & 0 \\ 1 & 4 & 0 & 0 \end{bmatrix}$

c. $A = \begin{bmatrix} 5 & 4 \\ 1 & 1 \end{bmatrix}$ **d.** $A = \begin{bmatrix} -2 & 1 & 0 \\ 6 & -3 & 1 \\ 8 & -4 & 3 \end{bmatrix}$

14. a. $A = \begin{bmatrix} 9 & -3 \\ -4 & 2 \\ 1 & 1 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 3 & -3 & 1 & 1 \\ 6 & -6 & 0 & 2 \\ 9 & -9 & 1 & 3 \end{bmatrix}$

c. $A = \begin{bmatrix} 3 & -9 \\ -1 & 3 \end{bmatrix}$ **d.** $A = \begin{bmatrix} 2 & 3 & 8 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$

In Exercises 15–16, describe in words the inverse of the given one-to-one operator.

15. a. The reflection about the x-axis on R^2.

b. The rotation about the origin through an angle of $\pi/4$ on R^2.

16. a. The reflection about the yz-plane in R^3.

b. The rotation through an angle of $-18°$ about the z-axis in R^3.

In Exercises 17–18, use matrix inversion to confirm the stated result in R^2.

17. a. The inverse transformation for a reflection about $y = x$ is a reflection about $y = x$.

b. The inverse transformation for a rotation about the origin is a rotation about the origin.

18. a. The inverse transformation for a reflections about a coordinate axis is a reflection about that axis.

b. The inverse transformation for a reflection about the origin is a reflection about the origin.

19. Let $T : P_1 \to R^2$ be the function defined by the formula

$$T(p(x)) = (p(0), p(1))$$

a. Find $T(1 - 2x)$.

b. Show that T is a linear transformation.

c. Show that T is one-to-one.

d. Find $T^{-1}(2, 3)$, and sketch its graph.

20. In each part, determine whether $T : R^n \to R^n$ is one-to-one; if so, find $T^{-1}(x_1, x_2, \ldots, x_n)$.

a. $T(x_1, x_2, \ldots, x_n) = (0, x_1, x_2, \ldots, x_{n-1})$

b. $T(x_1, x_2, \ldots, x_n) = (x_n, x_{n-1}, \ldots, x_2, x_1)$

c. $T(x_1, x_2, \ldots, x_n) = (x_2, x_3, \ldots, x_n, x_1)$

21. Let $T : R^n \to R^n$ be the linear operator defined by the formula

$$T(x_1, x_2, \ldots, x_n) = (a_1x_1, a_2x_2, \ldots, a_nx_n)$$

where $a_1, \ldots, a_n$ are constants.

a. Under what conditions will T have an inverse?

b. Assuming that the conditions determined in part (a) are satisfied, find a formula for $T^{-1}(x_1, x_2, \ldots, x_n)$.

22. Let $T_A : R^4 \to R^2$ be multiplication by the matrix

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

Find parametric equations for the set of vectors that map into the vector $(1, 1)$, if any.

In Exercises 23–24, compute $(T_2 \circ T_1)(x, y)$.

23. $T_1(x, y) = (2x, 3y)$, $T_2(x, y) = (x - y, x + y)$

24. $T_1(x, y) = (2x, -3y, x + y)$, $T_2(x, y, z) = (x - y, y + z)$

25. Suppose that the linear transformations $T_1 : P_2 \to P_2$ and $T_2 : P_2 \to P_3$ are given by the formulas $T_1(p(x)) = p(x + 1)$ and $T_2(p(x)) = xp(x)$. Find $(T_2 \circ T_1)(a_0 + a_1x + a_2x^2)$.

26. Let $T_1 : P_n \to P_n$ and $T_2 : P_n \to P_n$ be the linear operators given by $T_1(p(x)) = p(x - 1)$ and $T_2(p(x)) = p(x + 1)$. Find $(T_1 \circ T_2)(p(x))$ and $(T_2 \circ T_1)(p(x))$.

27. Let $T_1 : M_{22} \to R$ and $T_2 : M_{22} \to M_{22}$ be the linear transformations given by $T_1(A) = \mathrm{tr}(A)$ and $T_2(A) = A^T$.

a. Find $(T_1 \circ T_2)(A)$, where $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

b. Can you find $(T_2 \circ T_1)(A)$? Explain.

28. Rework Exercise 27 if $T_1 : M_{22} \to M_{22}$ and $T_2 : M_{22} \to M_{22}$ are the linear transformations, $T_1(A) = kA$ and $T_2(A) = A^T$, where k is a scalar.

In Exercises 29–30, compute $(T_3 \circ T_2 \circ T_1)(x, y)$.

29. $T_1(x, y) = (-2y, 3x, x - 2y)$, $T_2(x, y, z) = (y, z, x)$, $T_3(x, y, z) = (x + z, y - z)$

30. $T_1(x,y) = (x+y, y, -x)$, $T_2(x,y,z) = (0, x+y+z, 3y)$, $T_3(x,y,z) = (3x+2y, 4z-x-3y)$

31. Let $T_1 : P_2 \to P_3$ and $T_2 : P_3 \to P_3$ be the linear transformations given by the formulas

$$T_1(p(x)) = xp(x) \quad \text{and} \quad T_2(p(x)) = p(x+1)$$

a. Find formulas for

$$T_1^{-1}(p(x)), T_2^{-1}(p(x))$$

and

$$(T_1^{-1} \circ T_2^{-1})(p(x))$$

b. Verify that $(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}$.

32. Let $T_1 : R^2 \to R^2$ and $T_2 : R^2 \to R^2$ be the linear operators given by the formulas

$$T_1(x,y) = (x+y, x-y) \quad \text{and} \quad T_2(x,y) = (2x+y, x-2y)$$

a. Show that T_1 and T_2 are one-to-one.

b. Find formulas for

$$T_1^{-1}(x,y), \quad T_2^{-1}(x,y), \quad (T_2 \circ T_1)^{-1}(x,y)$$

c. Verify that $(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}$.

33. Let $T_1 : V \to V$ be the linear operator given by $T_1(\mathbf{v}) = 4\mathbf{v}$. Find a linear operator $T_2 : V \to V$ such that $T_1 \circ T_2 = I$ and $T_2 \circ T_1 = I$.

34. Let $T_1 : M_{22} \to P_1$ and $T_2 : P_1 \to R^3$ be the linear transformations given by

$$T_1\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a+b) + (c+d)x$$

and $T_2(a+bx) = (a, b, a)$.

a. Find the formula for $T_2 \circ T_1$.

b. Show that $T_2 \circ T_1$ is not one-to-one by finding distinct 2×2 matrices A and B such that

$$(T_2 \circ T_1)(A) = (T_2 \circ T_1)(B)$$

c. Show that $T_2 \circ T_1$ is not onto by finding a vector (a, b, c) in R^3 that is not in the range of $T_2 \circ T_1$.

35. Let $T : R^3 \to R^3$ be the orthogonal projection of R^3 onto the xy-plane. Show that $T \circ T = T$.

36. ***(Calculus required)*** Let V be the vector space $C^1[0,1]$ and let $T : V \to R$ be defined by

$$T(\mathbf{f}) = f(0) + 2f'(0) + 3f'(1)$$

Verify that T is a linear transformation. Determine whether T is one-to-one, and justify your conclusion.

37. ***(Calculus required)*** The Fundamental Theorem of Calculus implies that integration and differentiation reverse the actions of each other. Define a transformation $D : P_n \to P_{n-1}$ by $D(p(x)) = p'(x)$, and define $J : P_{n-1} \to P_n$ by

$$J(p(x)) = \int_0^x p(t)\,dt$$

a. Show that D and J are linear transformations.

b. Explain why J is not the inverse transformation of D.

c. Can the domains and/or codomains of D and J be restricted so they are inverse linear transformations?

38. ***(Calculus required)*** Let

$$D(\mathbf{f}) = f'(x) \quad \text{and} \quad J(\mathbf{f}) = \int_0^x f(t)\,dt$$

be the linear transformations in Examples 11 and 12 of Section 8.1. Find $(J \circ D)(\mathbf{f})$ for

a. $\mathbf{f}(x) = x^2 + 3x + 2$. **b.** $\mathbf{f}(x) = \sin x$.

39. ***(Calculus required)*** Let $J : P_1 \to R$ be the integration transformation $J(\mathbf{p}) = \int_{-1}^{1} p(x)dx$. Determine whether J is one-to-one. Justify your answer.

40. ***(Calculus required)*** Let $D : P_n \to P_{n-1}$ be the differentiation transformation $D(p(x)) = p'(x)$. Determine whether D is onto, and justify your answer.

41. Let A be an $n \times n$ matrix such that $\det(A) = 0$, and let $T : R^n \to R^n$ be multiplication by A.

a. What can you say about the range of the matrix operator T? Give an example that illustrates your conclusion.

b. What can you say about the number of vectors that T maps into $\mathbf{0}$?

42. Answer the questions in Exercise 41 in the case where $\det(A) \neq 0$.

43. a. Is a composition of one-to-one matrix transformations one-to-one? Justify your conclusion.

b. Can the composition of a one-to-one matrix transformation and a matrix transformation that is not one-to-one be one-to-one? Account for both possible orders of composition and justify your conclusion.

Working with Proofs

44. Prove: If there exists an onto linear transformation $T : V \to W$ then $\dim(V) \geq \dim(W)$.

45. Prove: If $T : V \to W$ is a one-to-one linear transformation, then $T^{-1} : R(T) \to V$ is a one-to-one linear transformation.

46. Use the definition of $T_3 \circ T_2 \circ T_1$ given by Formula (6) to prove that

a. $T_3 \circ T_2 \circ T_1$ is a linear transformation.

b. $T_3 \circ T_2 \circ T_1 = (T_3 \circ T_2) \circ T_1$.

c. $T_3 \circ T_2 \circ T_1 = T_3 \circ (T_2 \circ T_1)$.

47. Let V and W be finite-dimensional vector space and let $T : V \to W$ be a linear transformation. Prove:

a. If $\dim(V) > \dim(W)$, then T cannot be one-to-one.

b. If $\dim(V) < \dim(W)$, then T cannot be onto.

48. Add parts (t), (u), and (v) to Equivalence Theorem 8.2.4 by proving that each of those statements is equivalent to the invertibility of A.

True-False Exercises

TF. In parts **(a)–(j)** determine whether the statement is true or false, and justify your answer.

a. $T : V \to W$ is one-to-one if and only if $T(\mathbf{u}) \neq T(\mathbf{v})$ whenever $\mathbf{u} \neq \mathbf{v}$.

b. $T : V \to W$ is one-to-one if and only if for each vector $\mathbf{w}$ in the range of T there is exactly one vector $\mathbf{v}$ in V such that $T(\mathbf{v}) = \mathbf{w}$.

c. The inverse of a one-to-one linear transformation is a linear transformation.

d. If a linear transformation T has an inverse, then the kernel of T is the zero subspace.

e. If $T : R^2 \to R^2$ is the orthogonal projection onto the x-axis, then $T^{-1} : R^2 \to R^2$ maps each point on the x-axis onto a line that is perpendicular to the x-axis.

f. If $T_1 : U \to V$ and $T_2 : V \to W$ are linear transformations, and if T_1 is not one-to-one, then neither is $T_2 \circ T_1$.

g. If A is an $n \times n$ matrix and if the linear system $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution, then the range of the matrix operator is not R^n.

h. If T_A and T_B are matrix operators on R^n, then $T_A(T_B(\mathbf{x})) = T_B(T_A(\mathbf{x}))$ for every vector $\mathbf{x}$ in R^n.

i. The kernel of a matrix transformation $T_A : R^n \to R^m$ is the same as the null space of A.

j. If there is a nonzero vector in the kernel of the matrix operator $T_A : R^n \to R^n$, then this operator is not one-to-one.

Working with Technology

T1. Consider the matrix transformation $T_A : R^3 \to R^4$, where

$$A = \begin{bmatrix} .23 & -.02 & .67 \\ 1.12 & .10 & .44 \\ -.03 & .11 & .12 \\ .09 & -.68 & .83 \end{bmatrix}$$

Use Theorem 8.2.3 to determine whether T_A is one-to-one.

T2. Consider the matrix transformation $T_B : R^4 \to R^3$, where

$$B = \begin{bmatrix} .52 & .42 & .91 & -.05 \\ -.01 & 1.11 & .37 & .78 \\ .21 & .73 & -.32 & .24 \end{bmatrix}$$

Use Theorem 8.2.3 to determine whether T_B is onto.

8.3 Isomorphism

In this section we will establish a fundamental connection between real finite-dimensional vector spaces and the Euclidean space R^n. This connection is not only important theoretically, but it has practical applications in that is allows us to perform vector computations in certain general vector spaces by working with the vectors in R^n.

Isomorphism

Although many of the theorems in this text have been concerned exclusively with the vector space R^n, this is not as limiting as it might seem. We will show that the vector space R^n is the "mother" of all real n-dimensional vector spaces in the sense that every n-dimensional vector space must have the same algebraic structure as R^n even though its vectors may not be expressed as n-tuples. To explain what we mean by this, we will need the following definition.

Definition 1

A linear transformation $T : V \to W$ that is both one-to-one and onto is said to be an ***isomorphism***, and W is said to be ***isomorphic*** to V.

In the exercises we will ask you to show that if $T : V \to W$ is an isomorphism, then $T^{-1} : W \to V$ is also an isomorphism. Accordingly, we will usually say simply that ***V and W are isomorphic*** and that ***T is an isomorphism between V and W***.

The word *isomorphic* is derived from the Greek words *iso*, meaning "identical," and *morphe*, meaning "form." This terminology is appropriate because, as we will now explain, isomorphic vector spaces have the same "algebraic form," even though they may consist of different kinds of objects. For example, the following diagram illustrates an isomorphism between P_2 and R^3

$$c_0 + c_1x + c_2x^2 \;\underset{T^{-1}}{\overset{T}{\rightleftarrows}}\; (c_0, c_1, c_2)$$

Although the vectors on the two sides of the arrows are different kinds of objects, the vector operations on each side mirror those on the other side. For example, for scalar multiplication we have

$$k(c_0 + c_1x + c_2x^2) \;\underset{T^{-1}}{\overset{T}{\rightleftarrows}}\; k(c_0, c_1, c_2)$$

$$kc_0 + kc_1x + kc_2x^2 \;\underset{T^{-1}}{\overset{T}{\rightleftarrows}}\; (kc_0, kc_1, kc_2)$$

and for vector addition we have

$$(c_0 + c_1x + c_2x^2) + (d_0 + d_1x + d_2x^2) \;\underset{T^{-1}}{\overset{T}{\rightleftarrows}}\; (c_0, c_1, c_2) + (d_0, d_1, d_2)$$

$$(c_0 + d_0) + (c_1 + d_1)x + (c_2 + d_2)x^2 \;\underset{T^{-1}}{\overset{T}{\rightleftarrows}}\; (c_0 + d_0, c_1 + d_1, c_2 + d_2)$$

The following theorem, which is one of the most basic results in linear algebra, reveals the fundamental importance of the vector space R^n.

Theorem 8.3.1

Every real n-dimensional vector space is isomorphic to R^n.

Theorem 8.3.1 tells us that every real n-dimensional vector space differs from R^n only in notation; the algebraic structures of the two spaces are the same.

Proof Let V be a real n-dimensional vector space. To prove that V is isomorphic to R^n we must find a linear transformation $T: V \to R^n$ that is one-to-one and onto. For this purpose, let

$$S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$$

be any basis for V, let

$$\mathbf{u} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n \tag{1}$$

be the representation of a vector $\mathbf{u}$ in V as a linear combination of the basis vectors, and let $T: V \to R^n$ be the coordinate map

$$T(\mathbf{u}) = (\mathbf{u})_S = (k_1, k_2, \dots, k_n) \tag{2}$$

We will show that T is linear, one-to-one, and onto and hence is an isomorphism. To prove the linearity, let $\mathbf{u}$ and $\mathbf{v}$ be vectors in V, let c be a scalar, and let

$$\mathbf{u} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n \quad \text{and} \quad \mathbf{v} = d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_n\mathbf{v}_n \tag{3}$$

be the representations of $\mathbf{u}$ and $\mathbf{v}$ as linear combinations of the basis vectors. Then it follows from (3) that

$$\begin{aligned} T(c\mathbf{u}) &= T(ck_1\mathbf{v}_1 + ck_2\mathbf{v}_2 + \cdots + ck_n\mathbf{v}_n) \\ &= (ck_1, ck_2, \dots, ck_n) \\ &= c(k_1, k_2, \dots, k_n) = cT(\mathbf{u}) \end{aligned}$$

and that

$$\begin{aligned} T(\mathbf{u} + \mathbf{v}) &= T\big((k_1 + d_1)\mathbf{v}_1 + (k_2 + d_2)\mathbf{v}_2 + \cdots + (k_n + d_n)\mathbf{v}_n\big) \\ &= (k_1 + d_1, k_2 + d_2, \dots, k_n + d_n) \\ &= (k_1, k_2, \dots, k_n) + (d_1, d_2, \dots, d_n) \\ &= T(\mathbf{u}) + T(\mathbf{v}) \end{aligned}$$

which shows that T is linear. To show that T is one-to-one, we must show that if $\mathbf{u}$ and $\mathbf{v}$ are distinct vectors in V, then so are their images in R^n. But if $\mathbf{u} \neq \mathbf{v}$, and if the representations of these vectors in terms of the basis vectors are as in (3), then we must have $k_i \neq d_i$ for at least one i. Thus,

$$T(\mathbf{u}) = (k_1, k_2, \ldots, k_n) \neq (d_1, d_2, \ldots, d_n) = T(\mathbf{v})$$

which shows that $\mathbf{u}$ and $\mathbf{v}$ have distinct images under T. Finally, the transformation T is onto, for if

$$\mathbf{w} = (k_1, k_2, \ldots, k_n)$$

is any vector in R^n, then it follows from (2) that $\mathbf{w}$ is the image under T of the vector

$$\mathbf{u} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n \quad \blacksquare$$

Whereas Theorem 8.3.1 tells us, in general, that every real n-dimensional vector space is isomorphic to R^n, it is Formula (2) in its proof that tells us how to find isomorphisms.

Theorem 8.3.2

If S is an ordered basis for a vector space V, then the coordinate map

$$\mathbf{u} \xrightarrow{T} (\mathbf{u})_S$$

is an isomorphism between V and R^n.

Remark Recall that coordinate maps depend on the order in which the basis vectors are listed. Thus, Theorem 8.3.2 actually describes many possible isomorphisms, one for each of the $n!$ possible orders in which the basis vectors can be listed.

EXAMPLE 1 | The Natural Isomorphism Between P_{n-1} and R^n

It follows from Theorem 8.3.2 that the coordinate map

$$a_0 + a_1x + \cdots + a_{n-1}x^{n-1} \xrightarrow{T} (a_0, a_1, \ldots, a_{n-1})$$

defines an isomorphism between P_{n-1} and R^n. This is called the ***natural isomorphism*** between those vector spaces.

EXAMPLE 2 | The Natural Isomorphism Between M_{22} and R^4

It follows from Theorem 8.3.2 that the coordinate map

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \xrightarrow{T} (a, b, c, d)$$

defines an isomorphism between M_{22} and R^4. This is a special case of the isomorphism that maps an $m \times n$ matrix into its coordinate vector. We call this the ***natural isomorphism*** between M_{mn} and R^{mn}.

CALCULUS REQUIRED

EXAMPLE 3 | Differentiation by Matrix Multiplication

Consider the differentiation transformation $D: P_3 \to P_2$ on the vector space of polynomials of degree 3 or less. If we map P_3 and P_2 into R^4 and R^3, respectively, by the natural isomorphisms, then the transformation D produces a corresponding matrix transformation from R^4 to R^3. Specifically, the derivative transformation

$$a_0 + a_1x + a_2x^2 + a_3x^3 \xrightarrow{D} a_1 + 2a_2x + 3a_3x^2$$

produces the matrix transformation

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} a_1 \\ 2a_2 \\ 3a_3 \end{bmatrix}$$

Thus, for example, the derivative

$$\frac{d}{dx}(2 + x + 4x^2 - x^3) = 1 + 8x - 3x^2$$

can be calculated as the matrix product

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

This idea is useful for constructing numerical algorithms to calculate derivatives.

EXAMPLE 4 | Working with Isomorphisms

Use the natural isomorphism between P_5 and R^6 to determine whether the following polynomials are linearly independent.

$$\begin{aligned} \mathbf{p}_1 &= 1 + 2x - 3x^2 + 4x^3 + x^5 \\ \mathbf{p}_2 &= 1 + 3x - 4x^2 + 6x^3 + 5x^4 + 4x^5 \\ \mathbf{p}_3 &= 3 + 8x - 11x^2 - 16x^3 + 10x^4 + 9x^5 \end{aligned}$$

Solution We will convert this to a matrix problem by creating a matrix whose rows are the coordinate vectors of the polynomials under the natural isomorphism and then determine whether those rows are linearly independent using elementary row operations.

The matrix whose rows are the coordinate vectors of the polynomials under the natural isomorphism is

$$A = \begin{bmatrix} 1 & 2 & -3 & 4 & 0 & 1 \\ 1 & 3 & -4 & 6 & 5 & 4 \\ 3 & 8 & -11 & 16 & 10 & 9 \end{bmatrix}$$

We leave it for you to use elementary row operations to reduce this matrix to the row echelon form

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

This matrix has only two nonzero rows, so the row space of A is two-dimensional. This means that its row vectors are linearly dependent and hence so are the polynomials.

Inner Product Space Isomorphisms

In the case where V is a real n-dimensional inner product space, both V and R^n have, in addition to their algebraic structure, a geometric structure arising from their respective inner products. Thus, it is reasonable to inquire if there exists an isomorphism from V to R^n that preserves the geometric structure as well as the algebraic structure. For example, we would want orthogonal vectors in V to have orthogonal counterparts in R^n, and we would want orthonormal sets in V to correspond to orthonormal sets in R^n.

In order for an isomorphism to preserve geometric structure, it obviously has to preserve inner products, since notions of length, angle, and orthogonality are all based on the inner product. Thus, if V and W are inner product spaces, then we call an isomorphism $T: V \to W$ an ***inner product space isomorphism*** if

$$\langle T(\mathbf{u}), T(\mathbf{v}) \rangle = \langle \mathbf{u}, \mathbf{v} \rangle \quad \text{for all } \mathbf{u} \text{ and } \mathbf{v} \text{ in } V$$

Remark Keep in mind that the inner product on the left side of this equation is for W and that on the right is for V.

The following analog of Theorem 8.3.2 provides an important method for obtaining inner product space isomorphisms between real inner product spaces and Euclidean vector spaces.

Theorem 8.3.3

If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an ordered orthonormal basis for a real n-dimensional inner product space V, then the coordinate map

$$\mathbf{u} \xrightarrow{T} (\mathbf{u})_S$$

is an inner product space isomorphism between V and the vector space R^n with the Euclidean inner product.

EXAMPLE 5 | An Inner Product Space Isomorphism

We saw in Example 1 that the coordinate map

$$a_0 + a_1x + \cdots + a_{n-1}x^{n-1} \xrightarrow{T} (a_0, a_1, \ldots, a_{n-1})$$

with respect to the standard basis for P_{n-1} is an isomorphism between P_{n-1} and R^n. However, the standard basis is orthonormal with respect to the standard inner product on P_{n-1} (see Example 3 of Section 6.3), so it follows that T is actually an *inner product space isomorphism* with respect to the standard inner product on P_{n-1} and the Euclidean inner product on R^n. To verify that this is so, recall from Example 7 of Section 6.1 that the standard inner product on P_{n-1} of two vectors

$$\mathbf{p} = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} \quad \text{and} \quad \mathbf{q} = b_0 + b_1x + \cdots + b_{n-1}x^{n-1}$$

is

$$\langle \mathbf{p}, \mathbf{q} \rangle = a_0b_0 + a_1b_1 + \cdots + a_{n-1}b_{n-1}$$

But this is exactly the Euclidean inner product on R^n of the n-tuples

$$(a_0, a_1, \ldots, a_{n-1}) \quad \text{and} \quad (b_0, b_1, \ldots, b_{n-1})$$

EXAMPLE 6 | A Notational Matter

Let R^n be the vector space of real n-tuples in comma-delimited form, let M_n be the vector space of real $n \times 1$ matrices, let R^n have the Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v}$, and let M_n have the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T\mathbf{v}$ in which $\mathbf{u}$ and $\mathbf{v}$ are expressed in column form. The mapping $T: R^n \to M_n$ defined by

$$(v_1, v_2, \ldots, v_n) \xrightarrow{T} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

is an inner product space isomorphism, so the distinction between the inner product space R^n and the inner product space M_n is essentially notational, a fact that we have used many times in this text.

Exercise Set 8.3

In Exercises 1–8, state whether the transformation is an isomorphism. No proof required.

1. $c_0 + c_1x \to (c_0 - c_1, c_1)$ from P_1 to R^2.

2. $(x, y) \to (x, y, 0)$ from R^2 to R^3.

3. $a + bx + cx^2 + dx^3 \to \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ from P_3 to M_{22}.

4. $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \to ad - bc$ from M_{22} to R.

5. $(a, b, c, d) \to a + bx + cx^2 + (d + 1)x^3$ from R^4 to P_3.

6. $A \to A^T$ from M_{nn} to M_{nn}.

7. $c_1 \sin x + c_2 \cos x \to (c_1, c_2)$ from the subspace of $C(-\infty, \infty)$ spanned by $S = \{\sin x, \cos x\}$ to R^2.

8. The map $(u_1, u_2, \ldots, u_n, \ldots) \to (0, u_1, u_2, \ldots, u_n, \ldots)$ from R^∞ to R^∞.

9. **a.** Find an isomorphism between the vector space of all 3×3 symmetric matrices and R^6.

b. Find two different isomorphisms between the vector space of all 2×2 matrices and R^4.

10. **a.** Find an isomorphism between the vector space of all polynomials of degree at most 3 such that $p(0) = 0$ and R^3.

b. Find an isomorphism between the vector spaces span$\{1, \sin(x), \cos(x)\}$ and R^3.

In Exercises 11–12, determine whether the matrix transformation $T_A: R^3 \to R^3$ is an isomorphism.

11. $A = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & 2 \\ -1 & 1 & 0 \end{bmatrix}$ **12.** $A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 2 \\ -1 & 1 & 0 \end{bmatrix}$

In Exercises 13–14, find the dimension n of the solution space W of $A\mathbf{x} = \mathbf{0}$, and then construct an isomorphism between W and R^n.

13. $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \end{bmatrix}$ **14.** $A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$

In Exercises 15–16, determine whether the transformation is an isomorphism from M_{22} to R^4.

15. $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \to \begin{bmatrix} a \\ a+b \\ a+b+c \\ a+b+c+d \end{bmatrix}$

16. $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \to \begin{bmatrix} a+b \\ a+b \\ a+b+c \\ a+b+c+d \end{bmatrix}$

17. Do you think that R^2 is isomorphic to the xy-plane in R^3? Justify your answer.

18. **a.** For what value or values of k, if any, is M_{mn} isomorphic to R^k?

b. For what value or values of k, if any, is M_{mn} isomorphic to P_k?

19. Let $T: P_2 \to M_{22}$ be the mapping

$$T(\mathbf{p}) = T(p(x)) = \begin{bmatrix} p(0) & p(1) \\ p(1) & p(0) \end{bmatrix}$$

Is this an isomorphism? Justify your answer.

20. Show that if M_{22} and P_3 have the standard inner products given in Examples 6 and 7 of Section 6.1, then the mapping

$$\begin{bmatrix} a_0 & a_1 \\ a_2 & a_3 \end{bmatrix} \to a_0 + a_1x + a_2x^2 + a_3x^3$$

is an inner product space isomorphism between those spaces.

21. ***(Calculus required)*** Devise a method for using matrix multiplication to differentiate functions in the vector space span$\{1, \sin(x), \cos(x), \sin(2x), \cos(2x)\}$. Use your method to find the derivative of $3 - 4\sin(x) + \sin(2x) + 5\cos(2x)$.

Working with Proofs

22. Prove that if $T: V \to W$ is an isomorphism, then so is $T^{-1}: W \to V$.

23. Prove that if U, V, and W are vector spaces such that U is isomorphic to V and V is isomorphic to W, then U is isomorphic to W.

24. Use the result in Exercise 22 to prove that any two real finite-dimensional vector spaces with the same dimension are isomorphic to one another.

25. Prove that an inner product space isomorphism preserves angles and distances—that is, the angle between $\mathbf{u}$ and $\mathbf{v}$ in V is equal to the angle between $T(\mathbf{u})$ and $T(\mathbf{v})$ in W, and $\|\mathbf{u} - \mathbf{v}\|_V = \|T(\mathbf{u}) - T(\mathbf{v})\|_W$.

26. Prove that an inner product space isomorphism maps orthonormal sets into orthonormal sets.

True-False Exercises

TF. In parts **(a)–(f)** determine whether the statement is true or false, and justify your answer.

a. The vector spaces R^2 and P_2 are isomorphic.

b. If the kernel of a linear transformation $T: P_3 \to P_3$ is $\{\mathbf{0}\}$, then T is an isomorphism.

c. Every linear transformation from M_{33} to P_9 is an isomorphism.

d. There is a subspace of M_{23} that is isomorphic to R^4.

e. Isomorphic finite-dimensional vector spaces must have the same number of basis vectors.

f. R^n is isomorphic to a subspace of R^{n+1}.

8.4 Matrices for General Linear Transformations

In this section we will show that a general linear transformation from any n-dimensional vector space V to any m-dimensional vector space W can be performed using an appropriate matrix transformation from R^n to R^m. This idea is used in computer computations since computers are well suited for performing matrix computations.

Matrices of Linear Transformations

Suppose that V is an n-dimensional vector space, that W is an m-dimensional vector space, and that $T: V \to W$ is a linear transformation. Suppose further that B is a basis for V, that B' is a basis for W, and that for each vector $\mathbf{x}$ in V, the coordinate vectors for $\mathbf{x}$ and $T(\mathbf{x})$ are $[\mathbf{x}]_B$ and $[T(\mathbf{x})]_{B'}$, respectively (**Figure 8.4.1**).

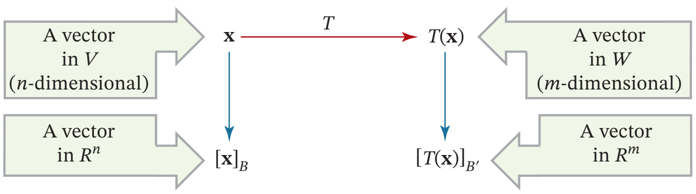


FIGURE 8.4.1

It will be our goal to find an $m \times n$ matrix A such that multiplication by A maps the vector $[\mathbf{x}]_B$ into the vector $[T(\mathbf{x})]_{B'}$ for each $\mathbf{x}$ in V (**Figure 8.4.2*a***). If we can do so, then, as illustrated in **Figure 8.4.2*b***, we will be able to execute the linear transformation T by using matrix multiplication and the following *indirect* procedure:

Finding $T(\mathbf{x})$ Indirectly

Step 1. Compute the coordinate vector $[\mathbf{x}]_B$.

Step 2. Multiply $[\mathbf{x}]_B$ on the left by A to produce $[T(\mathbf{x})]_{B'}$.

Step 3. Reconstruct $T(\mathbf{x})$ from its coordinate vector $[T(\mathbf{x})]_{B'}$.

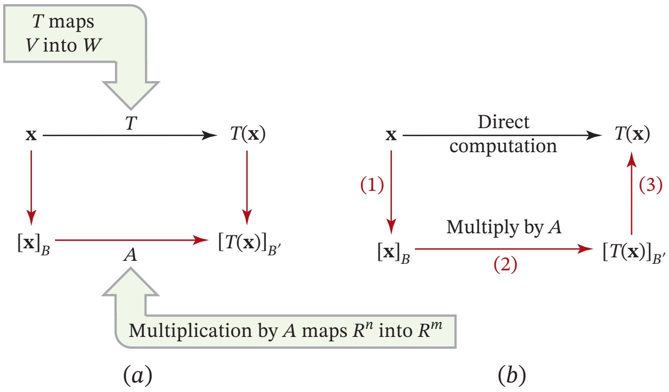


FIGURE 8.4.2

The key to executing this plan is to find an $m \times n$ matrix A with the property that

$$A[\mathbf{x}]_B = [T(\mathbf{x})]_{B'} \tag{1}$$

For this purpose, let $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a basis for the n-dimensional space V and $B' = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ a basis for the m-dimensional space W. Since Equation (1) must hold for all vectors in V, it must hold, in particular, for the basis vectors in B; that is,

$$A[\mathbf{u}_1]_B = [T(\mathbf{u}_1)]_{B'}, \quad A[\mathbf{u}_2]_B = [T(\mathbf{u}_2)]_{B'}, \ldots, \quad A[\mathbf{u}_n]_B = [T(\mathbf{u}_n)]_{B'} \tag{2}$$

But

$$[\mathbf{u}_1]_B = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad [\mathbf{u}_2]_B = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \quad [\mathbf{u}_n]_B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

so

$$A[\mathbf{u}_1]_B = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$$

$$A[\mathbf{u}_2]_B = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$A[\mathbf{u}_n]_B = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

Substituting these results into (2) yields

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} = [T(\mathbf{u}_1)]_{B'}, \quad \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} = [T(\mathbf{u}_2)]_{B'}, \ldots, \quad \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = [T(\mathbf{u}_n)]_{B'}$$

which shows that the successive columns of A are the coordinate vectors of

$$T(\mathbf{u}_1), T(\mathbf{u}_2), \ldots, T(\mathbf{u}_n)$$

with respect to the basis B'. Thus, the matrix A that completes the link in Figure 8.4.2*a* is

$$A = \big[[T(\mathbf{u}_1)]_{B'} \mid [T(\mathbf{u}_2)]_{B'} \mid \cdots \mid [T(\mathbf{u}_n)]_{B'}\big] \tag{3}$$

We will call this the ***matrix for T relative to the bases B and B′*** and will denote it by the symbol $[T]_{B',B}$. Using this notation, Formula (3) can be written as

$$\boxed{[T]_{B',B} = \big[[T(\mathbf{u}_1)]_{B'} \mid [T(\mathbf{u}_2)]_{B'} \mid \cdots \mid [T(\mathbf{u}_n)]_{B'}\big]} \tag{4}$$

and from (1), this matrix has the property

$$\boxed{[T]_{B',B}[\mathbf{x}]_B = [T(\mathbf{x})]_{B'}} \tag{5}$$

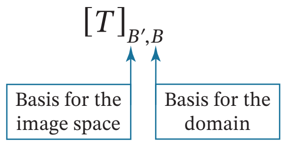


FIGURE 8.4.3

Remark Observe that in the notation $[T]_{B',B}$ the right subscript is a basis for the domain of T, and the left subscript is a basis for the image space of T (**Figure 8.4.3**). Moreover, observe how the subscript B seems to "cancel out" in Formula (5) (**Figure 8.4.4**).

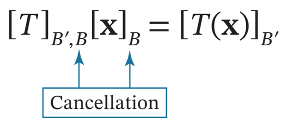


FIGURE 8.4.4

We leave it as an exercise to show that in the special case where $T_C: R^n \to R^m$ is multiplication by C, and where S and S' are the *standard bases* for R^n and R^m, respectively, then

$$[T_C]_{S',S} = C \tag{6}$$

EXAMPLE 1 | Matrix for a Linear Transformation

Let $T: P_1 \to P_2$ be the linear transformation defined by

$$T(p(x)) = xp(x)$$

Find the matrix for T with respect to the standard bases

$$B = \{\mathbf{u}_1, \mathbf{u}_2\} \quad \text{and} \quad B' = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$$

That is,

$$\mathbf{u}_1 = 1, \quad \mathbf{u}_2 = x; \qquad \mathbf{v}_1 = 1, \quad \mathbf{v}_2 = x, \quad \mathbf{v}_3 = x^2$$

Solution From the given formula for T we obtain

$$T(\mathbf{u}_1) = T(1) = (x)(1) = x$$
$$T(\mathbf{u}_2) = T(x) = (x)(x) = x^2$$

By inspection, the coordinate vectors for $T(\mathbf{u}_1)$ and $T(\mathbf{u}_2)$ relative to B' are

$$[T(\mathbf{u}_1)]_{B'} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad [T(\mathbf{u}_2)]_{B'} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Thus, the matrix for T with respect to B and B' is

$$[T]_{B',B} = \big[[T(\mathbf{u}_1)]_{B'} \mid [T(\mathbf{u}_2)]_{B'}\big] = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

EXAMPLE 2 | The Three-Step Procedure

Let $T : P_1 \to P_2$ be the linear transformation in Example 1, and use the three-step procedure illustrated in the following figure to perform the computation

$$T(a+bx) = x(a+bx) = ax + bx^2$$

$$\begin{tikzcd}
\mathbf{x} \arrow[r, "\text{Direct}", "\text{computation}"'] \arrow[d, "(1)"'] & T(\mathbf{x}) \\
{[\mathbf{x}]_B} \arrow[r, "\text{Multiply by } {[T]_{B',B}}", "(2)"'] & {[T(\mathbf{x})]_{B'}} \arrow[u, "(3)"']
\end{tikzcd}$$

Solution

Step 1. The coordinate vector for $\mathbf{x} = a + bx$ relative to the basis $B = \{1, x\}$ is

$$[\mathbf{x}]_B = \begin{bmatrix} a \\ b \end{bmatrix}$$

Step 2. Multiplying $[\mathbf{x}]_B$ by the matrix $[T]_{B',B}$ found in Example 1 we obtain

$$[T]_{B',B}[\mathbf{x}]_B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ a \\ b \end{bmatrix} = [T(\mathbf{x})]_{B'}$$

Step 3. Reconstructing $T(\mathbf{x}) = T(a+bx)$ from $[T(\mathbf{x})]_{B'}$ we obtain

$$T(a+bx) = 0 + ax + bx^2 = ax + bx^2$$

EXAMPLE 3 | Matrix for a Linear Transformation

Let $T : R^2 \to R^3$ be the linear transformation defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_2 \\ -5x_1 + 13x_2 \\ -7x_1 + 16x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -5 & 13 \\ -7 & 16 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Find the matrix for the transformation T with respect to the bases $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ for R^2 and $B' = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for R^3, where

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}; \qquad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

Solution From the formula for T,

$$T(\mathbf{u}_1) = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}, \quad T(\mathbf{u}_2) = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$$

Expressing these vectors as linear combinations of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, we obtain (verify)

$$T(\mathbf{u}_1) = \mathbf{v}_1 - 2\mathbf{v}_3, \quad T(\mathbf{u}_2) = 3\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3$$

Thus,

$$[T(\mathbf{u}_1)]_{B'} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}, \quad [T(\mathbf{u}_2)]_{B'} = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$$

so

$$[T]_{B',B} = [[T(\mathbf{u}_1)]_{B'} \mid [T(\mathbf{u}_2)]_{B'}] = \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ -2 & -1 \end{bmatrix}$$

Remark Example 3 illustrates that a fixed linear transformation generally has multiple representations, each depending on the bases chosen. In this case the matrices

$$[T]_{S',S} = \begin{bmatrix} 0 & 1 \\ -5 & 13 \\ -7 & 16 \end{bmatrix} \quad \text{and} \quad [T]_{B',B} = \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ -2 & -1 \end{bmatrix}$$

both represent the transformation T, the first relative to the standard bases, S and S', for R^2 and R^3, and the second relative to the bases B and B' stated in the example.

Matrices of Linear Operators

In the special case where $V = W$ (so that $T: V \to V$ is a linear operator), it is usual to take $B = B'$ when constructing a matrix for T. In this case the resulting matrix is called the ***matrix for T relative to the basis B*** and is usually denoted by $[T]_B$ rather than $[T]_{B,B}$. If $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$, then Formulas (4) and (5) become

$$\boxed{[T]_B = \big[[T(\mathbf{u}_1)]_B \mid [T(\mathbf{u}_2)]_B \mid \cdots \mid [T(\mathbf{u}_n)]_B\big]} \tag{7}$$

$$\boxed{[T]_B[\mathbf{x}]_B = [T(\mathbf{x})]_B} \tag{8}$$

Phrased informally, Formulas (7) and (8) state that *the matrix for T, when multiplied by the coordinate vector for $\mathbf{x}$, produces the coordinate vector for $T(\mathbf{x})$.*

We leave it for you to verify that if $T: R^n \to R^n$ is a matrix operator, say multiplication by A, and B is the standard basis for R^n, then Formula (7) simplifies to

$$[T_A]_B = A \tag{9}$$

Matrices of Identity Operators

Recall that the identity operator $I: V \to V$ maps each vector in a vector space V into itself, that is, $I(\mathbf{x}) = \mathbf{x}$ for every vector $\mathbf{x}$ in V. The following example shows that if V is n-dimensional, then the matrix for I relative to *any* basis B for V is the $n \times n$ identity matrix.

EXAMPLE 4 | Matrices of Identity Operators

If $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is a basis for an n-dimensional vector space V, and if $I: V \to V$ is the identity operator on V, then

$$I(\mathbf{u}_1) = \mathbf{u}_1, \quad I(\mathbf{u}_2) = \mathbf{u}_2, \ldots, \quad I(\mathbf{u}_n) = \mathbf{u}_n$$

Therefore,

$$[I]_B = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I$$

$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ [I(\mathbf{u}_1)]_B & [I(\mathbf{u}_2)]_B & [I(\mathbf{u}_n)]_B \end{array}$$

EXAMPLE 5 | Linear Operator on P_2

Let $T: P_2 \to P_2$ be the linear operator defined by

$$T(p(x)) = p(3x-5)$$

that is, $T(c_0 + c_1x + c_2x^2) = c_0 + c_1(3x-5) + c_2(3x-5)^2$.

(*a*) Find $[T]_B$ relative to the basis $B = \{1, x, x^2\}$.

(*b*) Use the indirect procedure to compute $T(1 + 2x + 3x^2)$.

(*c*) Check the result in (b) by computing $T(1 + 2x + 3x^2)$ directly.

Solution (*a*) From the formula for T,

$$T(1) = 1, \quad T(x) = 3x - 5, \quad T(x^2) = (3x-5)^2 = 9x^2 - 30x + 25$$

so

$$[T(1)]_B = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad [T(x)]_B = \begin{bmatrix} -5 \\ 3 \\ 0 \end{bmatrix}, \quad [T(x^2)]_B = \begin{bmatrix} 25 \\ -30 \\ 9 \end{bmatrix}$$

Thus,

$$[T]_B = \begin{bmatrix} 1 & -5 & 25 \\ 0 & 3 & -30 \\ 0 & 0 & 9 \end{bmatrix}$$

Solution (*b*)

Step 1. The coordinate vector for $\mathbf{p} = 1 + 2x + 3x^2$ relative to the basis $B = \{1, x, x^2\}$ is

$$[\mathbf{p}]_B = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Step 2. Multiplying $[\mathbf{p}]_B$ by the matrix $[T]_B$ found in part (a) we obtain

$$[T]_B[\mathbf{p}]_B = \begin{bmatrix} 1 & -5 & 25 \\ 0 & 3 & -30 \\ 0 & 0 & 9 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 66 \\ -84 \\ 27 \end{bmatrix} = [T(\mathbf{p})]_B$$

Step 3. Reconstructing $T(\mathbf{p}) = T(1 + 2x + 3x^2)$ from $[T(\mathbf{p})]_B$ we obtain

$$T(1 + 2x + 3x^2) = 66 - 84x + 27x^2$$

Solution (*c*) By direct computation,

$$\begin{aligned} T(1 + 2x + 3x^2) &= 1 + 2(3x-5) + 3(3x-5)^2 \\ &= 1 + 6x - 10 + 27x^2 - 90x + 75 \\ &= 66 - 84x + 27x^2 \end{aligned}$$

which agrees with the result in (b).

Matrices of Compositions and Inverse Transformations

We will conclude this section by mentioning two theorems without proof that are generalizations of earlier results.

Theorem 8.4.1

If $T_1: U \to V$ and $T_2: V \to W$ are linear transformations, and if B, B'', and B' are bases for U, V, and W, respectively, then

$$[T_2 \circ T_1]_{B',B} = [T_2]_{B',B''}[T_1]_{B'',B} \tag{10}$$

Theorem 8.4.2

If $T: V \to V$ is a linear operator, and if B is a basis for V, then the following are equivalent.

(*a*) T is one-to-one.

(*b*) $[T]_B$ is invertible.

Moreover, when these equivalent conditions hold,

$$[T^{-1}]_B = [T]_B^{-1} \tag{11}$$

Remark In (10), observe how the interior subscript B'' (the basis for the intermediate space V) seems to "cancel out," leaving only the bases for the domain and image space of the composition as subscripts (**Figure 8.4.5**). This "cancellation" of interior subscripts suggests the following extension of Formula (10) to compositions of three linear transformations (**Figure 8.4.6**):

$$\boxed{[T_3 \circ T_2 \circ T_1]_{B',B} = [T_3]_{B',B'''}[T_2]_{B''',B''}[T_1]_{B'',B}} \tag{12}$$

$$[T_2 \circ T_1]_{B',B} = [T_2]_{B',B''}[T_1]_{B'',B}$$

Cancellation

FIGURE 8.4.5

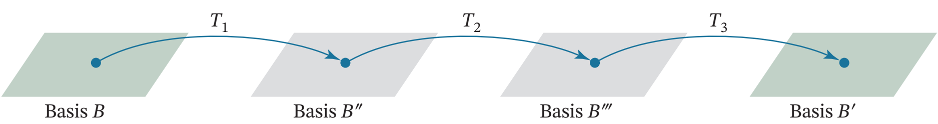


FIGURE 8.4.6

The following example illustrates Theorem 8.4.1.

EXAMPLE 6 | Composition

Let $T_1 : P_1 \to P_2$ be the linear transformation defined by

$$T_1(p(x)) = xp(x)$$

and let $T_2 : P_2 \to P_2$ be the linear operator defined by

$$T_2(p(x)) = p(3x - 5)$$

Then the composition $(T_2 \circ T_1) : P_1 \to P_2$ is given by

$$(T_2 \circ T_1)(p(x)) = T_2(T_1(p(x))) = T_2(xp(x)) = (3x - 5)p(3x - 5)$$

Thus, if $p(x) = c_0 + c_1 x$, then

$$\begin{aligned}(T_2 \circ T_1)(c_0 + c_1 x) &= (3x - 5)(c_0 + c_1(3x - 5)) \\ &= c_0(3x - 5) + c_1(3x - 5)^2\end{aligned} \tag{13}$$

In this example, P_1 plays the role of U in Theorem 8.4.1, and P_2 plays the roles of both V and W; thus we can take $B' = B''$ in (10) so that the formula simplifies to

$$[T_2 \circ T_1]_{B',B} = [T_2]_{B'}[T_1]_{B',B} \tag{14}$$

Let us choose $B = \{1, x\}$ to be the basis for P_1 and choose $B' = \{1, x, x^2\}$ to be the basis for P_2. We showed in Examples 1 and 5 that

$$[T_1]_{B',B} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad [T_2]_{B'} = \begin{bmatrix} 1 & -5 & 25 \\ 0 & 3 & -30 \\ 0 & 0 & 9 \end{bmatrix}$$

Thus, it follows from (14) that

$$[T_2 \circ T_1]_{B',B} = \begin{bmatrix} 1 & -5 & 25 \\ 0 & 3 & -30 \\ 0 & 0 & 9 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -5 & 25 \\ 3 & -30 \\ 0 & 9 \end{bmatrix} \tag{15}$$

As a check, we will calculate $[T_2 \circ T_1]_{B',B}$ directly from Formula (4). Since $B = \{1, x\}$, it follows from Formula (4) with $\mathbf{u}_1 = 1$ and $\mathbf{u}_2 = x$ that

$$[T_2 \circ T_1]_{B',B} = [[(T_2 \circ T_1)(1)]_{B'} \mid [(T_2 \circ T_1)(x)]_{B'}] \tag{16}$$

Using (13) yields

$$(T_2 \circ T_1)(1) = 3x - 5 \quad \text{and} \quad (T_2 \circ T_1)(x) = (3x-5)^2 = 9x^2 - 30x + 25$$

From this and the fact that $B' = \{1, x, x^2\}$, it follows that

$$[(T_2 \circ T_1)(1)]_{B'} = \begin{bmatrix} -5 \\ 3 \\ 0 \end{bmatrix} \quad \text{and} \quad [(T_2 \circ T_1)(x)]_{B'} = \begin{bmatrix} 25 \\ -30 \\ 9 \end{bmatrix}$$

Substituting in (16) yields

$$[T_2 \circ T_1]_{B',B} = \begin{bmatrix} -5 & 25 \\ 3 & -30 \\ 0 & 9 \end{bmatrix}$$

which agrees with (15).

Exercise Set 8.4

1. Let $T: P_2 \to P_3$ be the linear transformation defined by $T(p(x)) = xp(x)$.

a. Find the matrix for T relative to the standard bases

$$B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} \quad \text{and} \quad B' = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$$

where

$$\mathbf{u}_1 = 1, \quad \mathbf{u}_2 = x, \quad \mathbf{u}_3 = x^2$$
$$\mathbf{v}_1 = 1, \quad \mathbf{v}_2 = x, \quad \mathbf{v}_3 = x^2, \quad \mathbf{v}_4 = x^3$$

b. Verify that the matrix $[T]_{B',B}$ obtained in part (a) satisfies Formula (5) for every vector $\mathbf{x} = c_0 + c_1x + c_2x^2$ in P_2.

2. Let $T: P_2 \to P_1$ be the linear transformation defined by

$$T(a_0 + a_1x + a_2x^2) = (a_0 + a_1) - (2a_1 + 3a_2)x$$

a. Find the matrix for the linear transformation T relative to the standard bases $B = \{1, x, x^2\}$ and $B' = \{1, x\}$ for P_2 and P_1.

b. Verify that the matrix $[T]_{B',B}$ obtained in part (a) satisfies Formula (5) for every vector $\mathbf{x} = c_0 + c_1x + c_2x^2$ in P_2.

3. Let $T: P_2 \to P_2$ be the linear operator defined by

$$T(a_0 + a_1x + a_2x^2) = a_0 + a_1(x-1) + a_2(x-1)^2$$

a. Find the matrix for the linear transformation T relative to the standard basis $B = \{1, x, x^2\}$ for P_2.

b. Verify that the matrix $[T]_B$ obtained in part (a) satisfies Formula (8) for every vector $\mathbf{x} = a_0 + a_1x + a_2x^2$ in P_2.

4. Let $T: R^2 \to R^2$ be the linear operator defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 - x_2 \\ x_1 + x_2 \end{bmatrix}$$

and let $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ be the basis for which

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

a. Find $[T]_B$.

b. Verify that Formula (8) holds for every vector $\mathbf{x}$ in R^2.

5. Let $T: R^2 \to R^3$ be defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 \\ -x_1 \\ 0 \end{bmatrix}$$

a. Find the matrix $[T]_{B',B}$ relative to the bases $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $B' = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$$

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

b. Verify that Formula (5) holds for every vector in R^2.

6. Let $T: R^3 \to R^3$ be the linear operator defined by

$$T(x_1, x_2, x_3) = (x_1 - x_2, x_2 - x_1, x_1 - x_3)$$

a. Find the matrix for the linear transformation T with respect to the basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where

$$\mathbf{v}_1 = (1, 0, 1), \quad \mathbf{v}_2 = (0, 1, 1), \quad \mathbf{v}_3 = (1, 1, 0)$$

b. Verify that Formula (8) holds for every vector in R^3.

c. Is T one-to-one? If so, find the matrix of T^{-1} with respect to the basis B.

7. Let $T: P_2 \to P_2$ be the linear operator $T(p(x)) = p(2x+1)$, that is,

$$T(c_0 + c_1x + c_2x^2) = c_0 + c_1(2x+1) + c_2(2x+1)^2$$

a. Find $[T]_B$ with respect to the basis $B = \{1, x, x^2\}$.

b. Use the three-step procedure illustrated in Example 2 to compute $T(2 - 3x + 4x^2)$.

c. Check the result obtained in part (b) by computing $T(2 - 3x + 4x^2)$ directly.

8. Let $T: P_2 \to P_3$ be the linear transformation defined by $T(p(x)) = xp(x-3)$, that is,

$$T(c_0 + c_1x + c_2x^2) = x(c_0 + c_1(x-3) + c_2(x-3)^2)$$

a. Find $[T]_{B',B}$ relative to the bases $B = \{1, x, x^2\}$ and $B' = \{1, x, x^2, x^3\}$.

b. Use the three-step procedure illustrated in Example 2 to compute $T(1 + x - x^2)$.

c. Check the result obtained in part (b) by computing $T(1 + x - x^2)$ directly.

9. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$, and let $A = \begin{bmatrix} 1 & 3 \\ -2 & 5 \end{bmatrix}$ be the matrix for $T: R^2 \to R^2$ relative to the basis $B = \{\mathbf{v}_1, \mathbf{v}_2\}$.

a. Find $[T(\mathbf{v}_1)]_B$ and $[T(\mathbf{v}_2)]_B$.

b. Find $T(\mathbf{v}_1)$ and $T(\mathbf{v}_2)$.

c. Find a formula for $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)$.

d. Use the formula obtained in (c) to compute $T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$.

10. Let $A = \begin{bmatrix} 3 & -2 & 1 & 0 \\ 1 & 6 & 2 & 1 \\ -3 & 0 & 7 & 1 \end{bmatrix}$ be the matrix for $T: R^4 \to R^3$ relative to the bases $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ and $B' = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$, where

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 4 \\ -1 \\ 2 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 6 \\ 9 \\ 4 \\ 2 \end{bmatrix}$$

$$\mathbf{w}_1 = \begin{bmatrix} 0 \\ 8 \\ 8 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} -7 \\ 8 \\ 1 \end{bmatrix}, \quad \mathbf{w}_3 = \begin{bmatrix} -6 \\ 9 \\ 1 \end{bmatrix}$$

a. Find $[T(\mathbf{v}_1)]_{B'}$, $[T(\mathbf{v}_2)]_{B'}$, $[T(\mathbf{v}_3)]_{B'}$, and $[T(\mathbf{v}_4)]_{B'}$.

b. Find $T(\mathbf{v}_1)$, $T(\mathbf{v}_2)$, $T(\mathbf{v}_3)$, and $T(\mathbf{v}_4)$.

c. Find a formula for $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}\right)$.

d. Use the formula obtained in (c) to compute $T\left(\begin{bmatrix} 2 \\ 2 \\ 0 \\ 0 \end{bmatrix}\right)$.

11. Let $A = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 0 & 5 \\ 6 & -2 & 4 \end{bmatrix}$ be the matrix for $T: P_2 \to P_2$ with respect to the basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where $\mathbf{v}_1 = 3x + 3x^2$, $\mathbf{v}_2 = -1 + 3x + 2x^2$, $\mathbf{v}_3 = 3 + 7x + 2x^2$.

a. Find $[T(\mathbf{v}_1)]_B$, $[T(\mathbf{v}_2)]_B$, and $[T(\mathbf{v}_3)]_B$.

b. Find $T(\mathbf{v}_1)$, $T(\mathbf{v}_2)$, and $T(\mathbf{v}_3)$.

c. Find a formula for $T(a_0 + a_1x + a_2x^2)$.

d. Use the formula obtained in (c) to compute $T(1 + x^2)$.

12. Let $T_1: P_1 \to P_2$ be the linear transformation defined by

$$T_1(p(x)) = xp(x)$$

and let $T_2: P_2 \to P_2$ be the linear operator defined by

$$T_2(p(x)) = p(2x+1)$$

Let $B = \{1, x\}$ and $B' = \{1, x, x^2\}$ be the standard bases for P_1 and P_2.

a. Find $[T_2 \circ T_1]_{B',B}$, $[T_2]_{B'}$, and $[T_1]_{B',B}$.

b. State a formula relating the matrices in part (a).

c. Verify that the matrices in part (a) satisfy the formula you stated in part (b).

13. Let $T_1: P_1 \to P_2$ be the linear transformation defined by

$$T_1(c_0 + c_1x) = 2c_0 - 3c_1x$$

and let $T_2: P_2 \to P_3$ be the linear transformation defined by

$$T_2(c_0 + c_1x + c_2x^2) = 3c_0x + 3c_1x^2 + 3c_2x^3$$

Let $B = \{1, x\}$, $B'' = \{1, x, x^2\}$, and $B' = \{1, x, x^2, x^3\}$.

a. Find $[T_2 \circ T_1]_{B',B}$, $[T_2]_{B',B''}$, and $[T_1]_{B'',B}$.

b. State a formula relating the matrices in part (a).

c. Verify that the matrices in part (a) satisfy the formula you stated in part (b).

14. Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ be a basis for a vector space V. Find the matrix with respect to B for the linear operator $T: V \to V$ defined by $T(\mathbf{v}_1) = \mathbf{v}_2$, $T(\mathbf{v}_2) = \mathbf{v}_3$, $T(\mathbf{v}_3) = \mathbf{v}_4$, $T(\mathbf{v}_4) = \mathbf{v}_1$.

15. Let $T: P_2 \to M_{22}$ be the linear transformation defined by

$$T(\mathbf{p}) = \begin{bmatrix} p(0) & p(1) \\ p(-1) & p(0) \end{bmatrix}$$

let B be the standard basis for M_{22}, and let $B' = \{1, x, x^2\}$, $B'' = \{1, 1+x, 1+x^2\}$ be bases for P_2.

a. Find $[T]_{B,B'}$ and $[T]_{B,B''}$.

b. For the matrices obtained in part (a), find

$$T(2 + 2x + x^2)$$

using the three-step procedure illustrated in Example 2.

c. Check the results obtained in part (b) by computing $T(2 + 2x + x^2)$ directly.

16. Let $T: M_{22} \to R^2$ be the linear transformation given by

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a+b+c \\ d \end{bmatrix}$$

and let B be the standard basis for M_{22}, B' the standard basis for R^2, and

$$B'' = \left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}\right\}$$

a. Find $[T]_{B',B}$ and $[T]_{B'',B}$.

b. Compute $T\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\right)$ using the three-step procedure that was illustrated in Example 2 for both matrices found in part (a).

c. Check the results obtained in part (b) by computing $T\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\right)$ directly.

17. ***(Calculus required)*** Let $D: P_2 \to P_2$ be the differentiation operator $D(\mathbf{p}) = p'(x)$.

a. Find the matrix for D relative to the basis $B = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ for P_2 in which $\mathbf{p}_1 = 1$, $\mathbf{p}_2 = x$, $\mathbf{p}_3 = x^2$.

b. Use the matrix in part (a) to compute $D(6 - 6x + 24x^2)$.

18. ***(Calculus required)*** Let $D: P_2 \to P_2$ be the differentiation operator $D(\mathbf{p}) = p'(x)$.

a. Find the matrix for D relative to the basis $B = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ for P_2 in which $\mathbf{p}_1 = 2$, $\mathbf{p}_2 = 2 - 3x$, $\mathbf{p}_3 = 2 - 3x + 8x^2$.

b. Use the matrix in part (a) to compute $D(6 - 6x + 24x^2)$.

19. ***(Calculus required)*** Let V be the vector space of real-valued functions defined on the interval $(-\infty, \infty)$, and let $D: V \to V$ be the differentiation operator.

a. Find the matrix for D relative to the basis $B = \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ for V in which $\mathbf{f}_1 = 1$, $\mathbf{f}_2 = \sin x$, $\mathbf{f}_3 = \cos x$.

b. Use the matrix in part (a) to compute

$$D(2 + 3\sin x - 4\cos x)$$

20. Let V be a four-dimensional vector space with basis B, let W be a seven-dimensional vector space with basis B', and let $T: V \to W$ be a linear transformation. Identify the four vector spaces that contain the vectors at the corners of the accompanying diagram.

$$\begin{tikzcd}
\mathbf{x} \arrow[rr, "\text{Direct}", "\text{computation}"'] \arrow[d, "(1)"'] & & T(\mathbf{x}) \\
{[\mathbf{x}]_B} \arrow[rr, "\text{Multiply by } [T]_{B',B}", "(2)"'] & & {[T(\mathbf{x})]_{B'}} \arrow[u, "(3)"']
\end{tikzcd}$$

FIGURE Ex-20

21. In each part, fill in the missing part of the equation.

a. $[T_2 \circ T_1]_{B',B} = [T_2]\underline{\quad ? \quad}[T_1]_{B'',B}$

b. $[T_3 \circ T_2 \circ T_1]_{B',B} = [T_3]\underline{\quad ? \quad}[T_2]_{B''',B''}[T_1]_{B'',B}$

Working with Proofs

22. Prove: If $T: V \to W$ is the zero transformation, then the matrix for T with respect to any bases for V and W is a zero matrix.

23. Prove: If B and B' are the standard bases for R^n and R^m, respectively, then the matrix for a linear transformation $T: R^n \to R^m$ relative to the bases B and B' is the standard matrix for T.

True-False Exercises

TF. In parts **(a)–(e)** determine whether the statement is true or false, and justify your answer.

a. If the matrix of a linear transformation $T: V \to W$ relative to some bases of V and W is $\begin{bmatrix} 2 & 4 \\ 0 & 3 \end{bmatrix}$, then there is a nonzero vector $\mathbf{x}$ in V such that $T(\mathbf{x}) = 2\mathbf{x}$.

b. If the matrix of a linear transformation $T: V \to W$ relative to bases for V and W is $\begin{bmatrix} 2 & 4 \\ 0 & 3 \end{bmatrix}$, then there is a nonzero vector $\mathbf{x}$ in V such that $T(\mathbf{x}) = 4\mathbf{x}$.

c. If the matrix of a linear transformation $T: V \to W$ relative to certain bases for V and W is $\begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$, then T is one-to-one.

d. If $S: V \to V$ and $T: V \to V$ are linear operators and B is a basis for V, then the matrix of $S \circ T$ relative to B is $[T]_B[S]_B$.

e. If $T: V \to V$ is an invertible linear operator and B is a basis for V, then the matrix for T^{-1} relative to B is $[T]_B^{-1}$.

8.5 Similarity

The matrix for a linear operator $T: V \to V$ depends on the basis selected for V. One of the fundamental problems of linear algebra is to choose a basis for V that makes the matrix for T as simple as possible—a diagonal or a triangular matrix, for example. In this section we will study this problem.

Simple Matrices for Linear Operators

Standard bases do not necessarily produce the simplest matrices for linear operators. For example, consider the matrix operator $T: R^2 \to R^2$ whose matrix relative to the standard basis $B = \{\mathbf{e}_1, \mathbf{e}_2\}$ for R^2 is

$$[T]_B = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} \tag{1}$$

Let us compare this matrix to the matrix $[T]_{B'}$ for the same operator T but relative to the basis $B' = \{\mathbf{u}'_1, \mathbf{u}'_2\}$ for R^2 in which

$$\mathbf{u}'_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}'_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \tag{2}$$

Since

$$T(\mathbf{u}'_1) = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = 2\mathbf{u}'_1 \quad \text{and} \quad T(\mathbf{u}'_2) = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix} = 3\mathbf{u}'_2$$

it follows that

$$[T(\mathbf{u}'_1)]_{B'} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \quad \text{and} \quad [T(\mathbf{u}'_2)]_{B'} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

so the matrix for T relative to the basis B' is

$$[T]_{B'} = \big[T(\mathbf{u}'_1)_{B'} \mid T(\mathbf{u}'_2)_{B'}\big] = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

This matrix, being diagonal, has a simpler form than $[T]_B$ and conveys clearly that the operator T scales $\mathbf{u}'_1$ by a factor of 2 and $\mathbf{u}'_2$ by a factor of 3, information that is not immediately evident from $[T]_B$.

One of the major themes in more advanced linear algebra courses is to determine the "simplest possible form" that can be obtained for the matrix of a linear operator by choosing the basis appropriately. Sometimes it is possible to obtain a diagonal matrix (as above, for example), whereas other times one must settle for a triangular matrix or some other form. We will only be able to touch on this important topic in this text.

The problem of finding a basis that produces the simplest possible matrix for a linear operator $T: V \to V$ can be attacked by first finding a matrix for T relative to *any* basis, typically a standard basis, where applicable, and then changing the basis in a way that simplifies the matrix. Before pursuing this idea, it will be helpful to revisit some concepts about changing bases.

A New View of Transition Matrices

Recall from Formulas (9) and (10) of Section 4.7 that if $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ and $B' = \{\mathbf{u}'_1, \mathbf{u}'_2, \ldots, \mathbf{u}'_n\}$ are bases for a vector space V, then the transition matrices from B to B' and from B' to B are

$$P_{B\to B'} = \big[[\mathbf{u}_1]_{B'} \mid [\mathbf{u}_2]_{B'} \mid \cdots \mid [\mathbf{u}_n]_{B'}\big] \tag{3}$$

$$P_{B'\to B} = \big[[\mathbf{u}'_1]_B \mid [\mathbf{u}'_2]_B \mid \cdots \mid [\mathbf{u}'_n]_B\big] \tag{4}$$

where the matrices $P_{B\to B'}$ and $P_{B'\to B}$ are inverses of each other. We also showed in Formulas (11) and (12) of that section that if $\mathbf{v}$ is any vector in V, then

$$P_{B\to B'}[\mathbf{v}]_B = [\mathbf{v}]_{B'} \tag{5}$$

$$P_{B'\to B}[\mathbf{v}]_{B'} = [\mathbf{v}]_B \tag{6}$$

The following theorem shows that transition matrices in Formulas (3) and (4) can be viewed as matrices for identity operators.

Theorem 8.5.1

If B and B' are bases for a finite-dimensional vector space V, and if $I : V \to V$ is the identity operator on V, then

$$P_{B\to B'} = [I]_{B',B} \quad \text{and} \quad P_{B'\to B} = [I]_{B,B'}$$

Proof Suppose that $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ and $B' = \{\mathbf{u}'_1, \mathbf{u}'_2, \ldots, \mathbf{u}'_n\}$ are bases for V. Using the fact that $I(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v}$ in V, it follows from Formula (4) of Section 8.4 that

$$\begin{aligned}[I]_{B',B} &= \big[[I(\mathbf{u}_1)]_{B'} \mid [I(\mathbf{u}_2)]_{B'} \mid \cdots \mid [I(\mathbf{u}_n)]_{B'}\big] \\ &= \big[[\mathbf{u}_1]_{B'} \mid [\mathbf{u}_2]_{B'} \mid \cdots \mid [\mathbf{u}_n]_{B'}\big] \\ &= P_{B\to B'} \qquad \textbf{[Formula (3)]}\end{aligned}$$

The proof that $[I]_{B,B'} = P_{B'\to B}$ is similar. ■

Effect of Changing Bases on Matrices of Linear Operators

We are now ready to consider the main problem in this section.

Problem

If B and B' are two bases for a finite-dimensional vector space V, and if $T : V \to V$ is a linear operator, what relationship, if any, exists between the matrices $[T]_B$ and $[T]_{B'}$?

The answer to this question can be obtained by considering the composition of the three linear operators on V pictured in **Figure 8.5.1**.

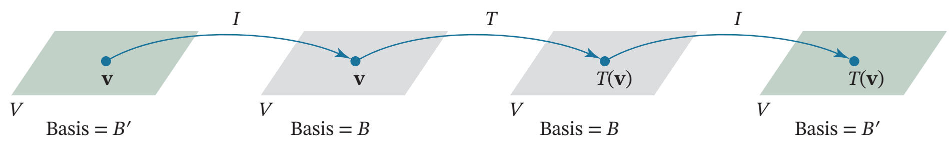


FIGURE 8.5.1

In this figure, $\mathbf{v}$ is first mapped into itself by the identity operator, then $\mathbf{v}$ is mapped into $T(\mathbf{v})$ by T, and then $T(\mathbf{v})$ is mapped into itself by the identity operator. All four vector spaces involved in the composition are the same (namely, V), but the bases for the spaces vary. Since the starting vector is $\mathbf{v}$ and the final vector is $T(\mathbf{v})$, the composition produces the same result as applying T directly; that is,

$$T = I \circ T \circ I \tag{7}$$

If, as illustrated in **Figure 8.5.1**, the first and last vector spaces are assigned the basis B' and the middle two spaces are assigned the basis B, then it follows from (7) and Formula (12) of Section 8.4 (with an appropriate adjustment to the names of the bases) that

$$[T]_{B',B'} = [I \circ T \circ I]_{B',B'} = [I]_{B',B}[T]_{B,B}[I]_{B,B'} \tag{8}$$

or, in simpler notation,

$$[T]_{B'} = [I]_{B',B}[T]_B[I]_{B,B'} \tag{9}$$

We can simplify this formula even further by using Theorem 8.5.1 to rewrite it as

$$\boxed{[T]_{B'} = P_{B\to B'}[T]_B P_{B'\to B}} \tag{10}$$

In summary, we have the following theorem.

Theorem 8.5.2

Let $T: V \to V$ be a linear operator on a finite-dimensional vector space V, and let B and B' be bases for V. Then

$$[T]_{B'} = P^{-1}[T]_B P \tag{11}$$

where $P = P_{B'\to B}$ and $P^{-1} = P_{B\to B'}$.

Warning When applying Theorem 8.5.2, it is easy to forget whether $P = P_{B'\to B}$ (correct) or $P = P_{B\to B'}$ (incorrect). It may help to use the diagram in **Figure 8.5.2** and observe that the *exterior* subscripts of the transition matrices match the subscript of the matrix they enclose.

$$[T]_{B'} = P_{B\to B'}\,[T]_B\,P_{B'\to B}$$

Exterior subscripts

FIGURE 8.5.2

In the terminology of Definition 1 of Section 5.2, Theorem 8.5.2 tells us that matrices representing the same linear operator relative to different bases must be similar. The following theorem, which we state without proof, shows that the converse of Theorem 8.5.2 is also true.

Theorem 8.5.3

If V is a finite-dimensional vector space, then two matrices A and B represent the same linear operator (but possibly with respect to different bases) if and only if they are similar. Moreover, if $B = P^{-1}AP$, then P is the transition matrix from the basis used for B to the basis used for A.

EXAMPLE 1 | Similar Matrices Represent the Same Linear Operator

We showed at the beginning of this section that the matrices

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

represent the same linear operator $T: R^2 \to R^2$ with respect to the appropriate bases. Verify that these matrices are similar by finding a matrix P for which $D = P^{-1}CP$.

Solution We need to find the transition matrix

$$P = P_{B'\to B} = [[\mathbf{u}_1']_B \mid [\mathbf{u}_2']_B]$$

where $B' = \{\mathbf{u}_1', \mathbf{u}_2'\}$ is the basis for R^2 given by (2) and $B = \{\mathbf{e}_1, \mathbf{e}_2\}$ is the standard basis for R^2. We see by inspection that

$$\begin{aligned} \mathbf{u}_1' &= \mathbf{e}_1 + \mathbf{e}_2 \\ \mathbf{u}_2' &= \mathbf{e}_1 + 2\mathbf{e}_2 \end{aligned}$$

from which it follows that

$$[\mathbf{u}_1']_B = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad [\mathbf{u}_2']_B = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Thus,

$$P = P_{B' \to B} = [[\mathbf{u}_1']_B \mid [\mathbf{u}_2']_B] = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

We leave it for you to verify that

$$P^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

and hence that

$$\underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}}_{D} = \underbrace{\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}}_{P^{-1}} \underbrace{\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}}_{P}$$

Similarity Invariants

Recall from Section 5.2 that a property of a square matrix is called a *similarity invariant* if that property is shared by all similar matrices. In **Table 1** of that section we listed the most important similarity invariants. Since we know from Theorem 8.5.3 that two matrices are similar if and only if they represent the same linear operator $T: V \to V$, it follows that if B and B' are bases for V, then every similarity invariant property of $[T]_B$ is also a similarity invariant property of $[T]_{B'}$. For example, for any two bases B and B' we must have

$$\det[T]_B = \det[T]_{B'}$$

It follows from this equation that the value of the determinant depends on T, but not on the particular basis that is used to represent T in matrix form. Thus, the determinant can be regarded as a property of the linear operator T, and we can *define* the ***determinant of the linear operator T*** to be

$$\det(T) = \det[T]_B \tag{12}$$

where B is *any* basis for V. **Table 1** lists the basic similarity invariants of a linear operator $T: V \to V$.

TABLE 1 Similarity Invariants

Property	Similarity
Determinant	$[T]_B$ and $P^{-1}[T]_BP$ have the same determinant.
Invertibility	$[T]_B$ is invertible if and only if $P^{-1}[T]_BP$ is invertible.
Rank	$[T]_B$ and $P^{-1}[T]_BP$ have the same rank.
Nullity	$[T]_B$ and $P^{-1}[T]_BP$ have the same nullity.
Trace	$[T]_B$ and $P^{-1}[T]_BP$ have the same trace.
Characteristic polynomial	$[T]_B$ and $P^{-1}[T]_BP$ have the same characteristic polynomial.
Eigenvalues	$[T]_B$ and $P^{-1}[T]_BP$ have the same eigenvalues.
Eigenspace dimension	If λ is an eigenvalue of $[T]_B$ and $P^{-1}[T]_BP$, then the eigenspace of $[T]_B$ corresponding to λ and the eigenspace of $P^{-1}[T]_BP$ corresponding to λ have the same dimension.

EXAMPLE 2 | Determinant of a Linear Operator

At the beginning of this section we showed that the matrices

$$[T] = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} \quad \text{and} \quad [T]_{B'} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

represent the same linear operator relative to different bases, the first relative to the standard basis $B = \{\mathbf{e}_1, \mathbf{e}_2\}$ for R^2 and the second relative to the basis $B' = \{\mathbf{u}_1', \mathbf{u}_2'\}$ for which

$$\mathbf{u}_1' = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2' = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

This means that $[T]$ and $[T]_{B'}$ must be similar matrices and hence must have the same similarity invariant properties. In particular, they must have the same determinant. The following computations confirm this.

$$\det[T] = \begin{vmatrix} 1 & 1 \\ -2 & 4 \end{vmatrix} = 6 \quad \text{and} \quad \det[T]_{B'} = \begin{vmatrix} 2 & 0 \\ 0 & 3 \end{vmatrix} = 6$$

EXAMPLE 3 | Eigenvalues of a Linear Operator

Find the eigenvalues of the linear operator $T: P_2 \to P_2$ defined by

$$T(a + bx + cx^2) = -2c + (a + 2b + c)x + (a + 3c)x^2$$

Solution Because eigenvalues are similarity invariants, we can find the eigenvalues of T by choosing any basis B for P_2 and computing the eigenvalues of the matrix $[T]_B$. We leave it for you to show that the matrix for T relative to the standard basis $B = \{1, x, x^2\}$ is

$$[T]_B = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$$

Thus, the eigenvalues of T are $\lambda = 1$ and $\lambda = 2$ (see Example 7 of Section 5.1).

Exercise Set 8.5

In Exercises 1–2, use a property from Table 1 to show that the matrices A and B are not similar.

1. **a.** $A = \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$

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

2. **a.** $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$

b. $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$

3. Let $T: R^2 \to R^2$ be a linear operator, and let B and B' be bases for R^2 for which

$$[T]_B = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad P_{B \to B'} = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}$$

Find the matrix for T relative to the basis B'.

4. Let $T: R^2 \to R^2$ be a linear operator, and let B and B' be bases for R^2 for which

$$[T]_B = \begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad P_{B' \to B} = \begin{bmatrix} 4 & 5 \\ 1 & -1 \end{bmatrix}$$

Find the matrix for T relative to the basis B'.

5. Let $T: R^2 \to R^2$ be a linear operator, and let B and B' be bases for R^2 for which

$$[T]_{B'} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad P_{B \to B'} = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}$$

Find the matrix for T relative to the basis B.

6. Let $T: R^2 \to R^2$ be a linear operator, and let B and B' be bases for R^2 for which

$$[T]_{B'} = \begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad P_{B' \to B} = \begin{bmatrix} 4 & 5 \\ 1 & -1 \end{bmatrix}$$

Find the matrix for T relative to the basis B.

In Exercises 7–14, find the matrix for T relative to the basis B, and use Theorem 8.5.2 to compute the matrix for T relative to the basis B'.

7. $T: R^2 \to R^2$ is defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 - 2x_2 \\ -x_2 \end{bmatrix}$$

and $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $B' = \{\mathbf{v}_1, \mathbf{v}_2\}$, where

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}; \quad \mathbf{v}_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 7 \\ 2 \end{bmatrix}$$

8. $T: R^2 \to R^2$ is defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 7x_2 \\ 3x_1 - 4x_2 \end{bmatrix}$$

and $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $B' = \{\mathbf{v}_1, \mathbf{v}_2\}$, where

$$\mathbf{u}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 4 \\ -1 \end{bmatrix}; \quad \mathbf{v}_1 = \begin{bmatrix} 18 \\ 8 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 10 \\ 5 \end{bmatrix}$$

9. $T: R^3 \to R^3$ is defined by

$$T(x_1, x_2, x_3) = (-2x_1 - x_2, x_1 + x_3, x_2)$$

B is the standard basis, and $B' = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where

$$\mathbf{v}_1 = (-2, 1, 0), \quad \mathbf{v}_2 = (-1, 0, 1), \quad \mathbf{v}_3 = (0, 1, 0)$$

10. $T: R^3 \to R^3$ is defined by

$$T(x_1, x_2, x_3) = (x_1 + 2x_2 - x_3, -x_2, x_1 + 7x_3)$$

B is the standard basis, and $B' = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where

$$\mathbf{v}_1 = (1, 0, 0), \quad \mathbf{v}_2 = (1, 1, 0), \quad \mathbf{v}_3 = (1, 1, 1)$$

11. $T: R^2 \to R^2$ is the rotation about the origin through an angle of 45°, B is the standard basis, and $B' = \{\mathbf{v}_1, \mathbf{v}_2\}$, where

$$\mathbf{v}_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \quad \mathbf{v}_2 = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

12. $T: R^2 \to R^2$ is the shear in the x-direction by a positive factor k, B is the standard basis, and $B' = \{\mathbf{v}_1, \mathbf{v}_2\}$, where

$$\mathbf{v}_1 = (k, 1), \quad \mathbf{v}_2 = (1, 0)$$

13. $T: P_1 \to P_1$ is defined by

$$T(a_0 + a_1x) = -a_0 + (a_0 + a_1)x$$

B is the standard basis for P_1, and $B' = \{\mathbf{q}_1, \mathbf{q}_2\}$, where

$$\mathbf{q}_1 = x + 1, \quad \mathbf{q}_2 = x - 1$$

14. $T: P_1 \to P_1$ is defined by $T(a_0 + a_1x) = a_0 + a_1(x + 1)$, and $B = \{\mathbf{p}_1, \mathbf{p}_2\}$ and $B' = \{\mathbf{q}_1, \mathbf{q}_2\}$, where

$$\mathbf{p}_1 = 6 + 3x, \quad \mathbf{p}_2 = 10 + 2x; \quad \mathbf{q}_1 = 2, \quad \mathbf{q}_2 = 3 + 2x$$

15. Let $T: P_2 \to P_2$ be defined by

$$T(a_0 + a_1x + a_2x^2) = (5a_0 + 6a_1 + 2a_2) - (a_1 + 8a_2)x + (a_0 - 2a_2)x^2$$

a. Find the eigenvalues of T.

b. Find bases for the eigenspaces of T.

16. Let $T: M_{22} \to M_{22}$ be defined by

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} 2c & a + c \\ b - 2c & d \end{bmatrix}$$

a. Find the eigenvalues of T.

b. Find bases for the eigenspaces of T.

17. Since the standard basis for R^n is so simple, why would one want to represent a linear operator on R^n in another basis?

18. Find two nonzero 2×2 matrices (different from those in Exercise 1) that are not similar, and explain why they are not.

In Exercises 19–21, find the determinant and the eigenvalues of the linear operator T.

19. $T: R^2 \to R^2$, where $T(x_1, x_2) = (3x_1 - 4x_2, -x_1 + 7x_2)$

20. $T: R^3 \to R^3$, where $T(x_1, x_2, x_3) = (x_1 - x_2, x_2 - x_3, x_3 - x_1)$

21. $T: P_2 \to P_2$, where $T(p(x)) = p(x - 1)$

22. Let $T: P_4 \to P_4$ be the linear operator given by the formula $T(p(x)) = p(2x + 1)$.

a. Find a matrix for T relative to some convenient basis, and then use it to find the rank and nullity of T.

b. Use the result in part (a) to determine whether T is one-to-one.

Working with Proofs

23. Complete the proof below by justifying each step.

Hypothesis: A and B are similar matrices.

Conclusion: A and B have the same characteristic polynomial.

Proof:
$$\begin{aligned}
&(1)\quad \det(\lambda I - B) = \det(\lambda I - P^{-1}AP) \\
&(2)\quad = \det(\lambda P^{-1}P - P^{-1}AP) \\
&(3)\quad = \det(P^{-1}(\lambda I - A)P) \\
&(4)\quad = \det(P^{-1})\det(\lambda I - A)\det(P) \\
&(5)\quad = \det(P^{-1})\det(P)\det(\lambda I - A) \\
&(6)\quad = \det(\lambda I - A)
\end{aligned}$$

24. If A and B are similar matrices, say $B = P^{-1}AP$, then it follows from Exercise 23 that A and B have the same eigenvalues. Suppose that λ is one of the common eigenvalues and $\mathbf{x}$ is a corresponding eigenvector of A. See if you can find an eigenvector of B corresponding to λ (expressed in terms of λ, $\mathbf{x}$, and P).

In Exercises 25–28, prove that the stated property is a similarity invariant.

25. Trace

26. Rank

27. Nullity

28. Invertibility

29. Let λ be an eigenvalue of a linear operator $T : V \to V$. Prove that the eigenvectors of T corresponding to λ are the nonzero vectors in the kernel of $\lambda I - T$.

30. a. Prove that if A and B are similar matrices, then A^2 and B^2 are also similar.

b. If A^2 and B^2 are similar, must A and B be similar? Explain.

31. Let C and D be $m \times n$ matrices, and let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for a vector space V. Prove that if $C[\mathbf{x}]_B = D[\mathbf{x}]_B$ for all $\mathbf{x}$ in V, then $C = D$.

True-False Exercises

TF. In parts **(a)–(h)** determine whether the statement is true or false, and justify your answer.

a. A matrix cannot be similar to itself.

b. If A is similar to B, and B is similar to C, then A is similar to C.

c. If A and B are similar and B is singular, then A is singular.

d. If A and B are invertible and similar, then A^{-1} and B^{-1} are similar.

e. If $T_1 : R^n \to R^n$ and $T_2 : R^n \to R^n$ are linear operators, and if $[T_1]_{B',B} = [T_2]_{B',B}$ with respect to two bases B and B' for R^n, then $T_1(\mathbf{x}) = T_2(\mathbf{x})$ for every vector $\mathbf{x}$ in R^n.

f. If $T_1 : R^n \to R^n$ is a linear operator, and if $[T_1]_B = [T_1]_{B'}$ with respect to two bases B and B' for R^n, then $B = B'$.

g. If $T : R^n \to R^n$ is a linear operator, and if $[T]_B = I_n$ with respect to some basis B for R^n, then T is the identity operator on R^n.

h. If $T : R^n \to R^n$ is a linear operator, and if $[T]_{B',B} = I_n$ with respect to two bases B and B' for R^n, then T is the identity operator on R^n.

Working with Technology

T1. Use the matrices A and P given below to construct a matrix $B = P^{-1}AP$ that is similar to A, and confirm, in accordance with Table 1, that A and B have the same determinant, trace, rank, characteristic equation, and eigenvalues.

$$A = \begin{bmatrix} -13 & -60 & -60 \\ 10 & 42 & 40 \\ -5 & -20 & -18 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 1 & -1 & 1 \\ 2 & -1 & -1 \\ -1 & -1 & 0 \end{bmatrix}$$

T2. Let $T : R^3 \to R^3$ be the linear transformation whose standard matrix is the matrix A in Exercise T1. Find a basis S for R^3 for which $[T]_S$ is diagonal.

8.6 Geometry of Matrix Operators

In applications such as computer graphics it is important to understand not only how linear operators on R^2 and R^3 affect individual vectors but also how they transform entire regions in two and three dimensions. In R^2, for example, one can get a sense of the effect of a linear operator on all regions in the plane by examining its effect on a unit square in the first quadrant. This will be our primary focus in this section. We will also continue our study of rotations by extending our work in Section 1.8 to three dimensions.

Computerized Transformations

Figure 8.6.1 shows a famous picture of Albert Einstein that has been transformed in various ways using matrix operators on R^2. The original image was scanned and then digitized to decompose it into a rectangular array of pixels. Those pixels were then transformed as follows:

- The program MATLAB was used to assign coordinates and a gray level to each pixel.
- The coordinates of the pixels were transformed by matrix multiplication.
- The pixels were then assigned their original gray levels to produce the transformed picture.

In computer games a perception of motion is created by using matrices to rapidly and repeatedly transform the arrays of pixels that form the visual images.

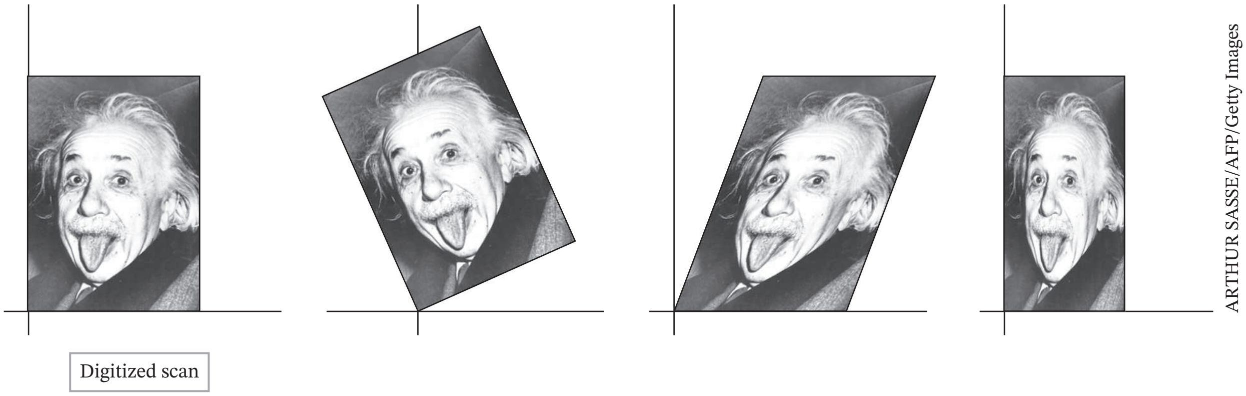


FIGURE 8.6.1

Images of Lines Under Matrix Operators

The effect of a matrix operator on R^2 can often be deduced by studying how it transforms the points that form the unit square. The following theorem, which we state without proof, shows that if an operator is invertible, then it maps each line segment in the unit square into the line segment connecting the images of its endpoints. In particular, the edges of the unit square get mapped into edges of the image (see **Figure 8.6.2** in which the edges of the unit square and the corresponding edges of its image have been numbered).

Theorem 8.6.1

If $T: R^2 \to R^2$ is multiplication by an invertible matrix, then:

(*a*) The image of a straight line is a straight line.

(*b*) The image of a line through the origin is a line through the origin.

(*c*) The images of parallel lines are parallel lines.

(*d*) The image of the line segment joining points P and Q is the line segment joining the images of P and Q.

(*e*) The images of three points lie on a line if and only if the points themselves lie on a line.

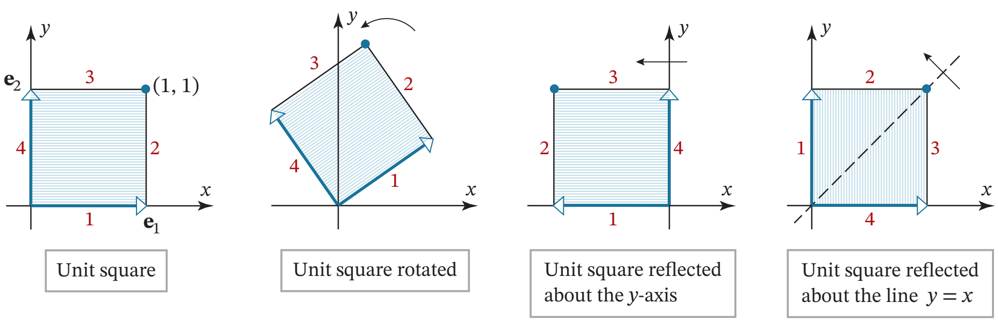


FIGURE 8.6.2

EXAMPLE 1 | Image of a Line

According to Theorem 8.6.1, the invertible matrix

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

maps the line $y = 2x + 1$ into another line. Find its equation.

Solution Let (x, y) be a point on the line $y = 2x + 1$, and let (x', y') be its image under multiplication by A. Then

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}^{-1}\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix}$$

so

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

Substituting these expressions in $y = 2x + 1$ yields

$$-2x' + 3y' = 2(x' - y') + 1$$

or, equivalently,

$$y' = \tfrac{4}{5}x' + \tfrac{1}{5}$$

EXAMPLE 2 | Transformation of the Unit Square

Sketch the image of the unit square under multiplication by the invertible matrix

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

Label the vertices of the image with their coordinates, and number the edges of the unit square and their corresponding images (as in Figure 8.6.2).

Solution Since

$$\begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix},$$

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

the image of the unit square is a parallelogram with vertices $(0, 0)$, $(0, 2)$, $(1, 1)$, and $(1, 3)$ (**Figure 8.6.3**).

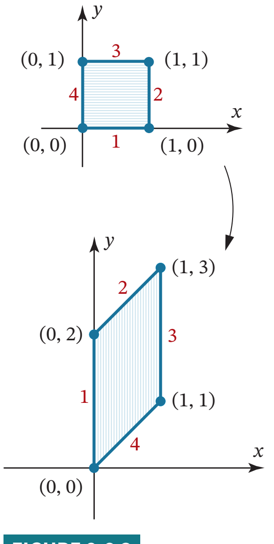
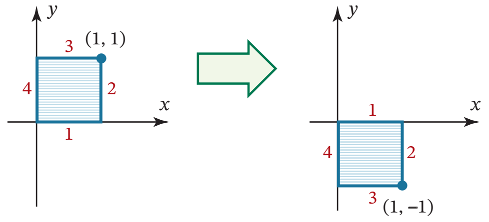
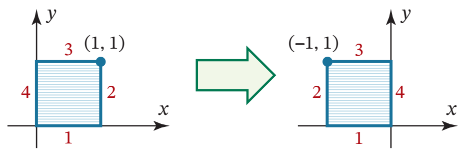
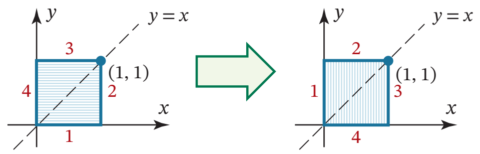


FIGURE 8.6.3

Reflections, Rotations, and Projections

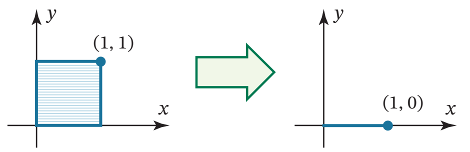
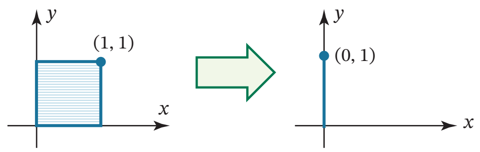
In Table 1 of Section 1.8 we obtained the standard matrices for the reflections about the x-axis, the y-axis, and the line $y = x$ in R^2. **Table 1** illustrates the effect of those transformations on the unit square.

TABLE 1

Operator	Standard Matrix	Effect on the Unit Square
Reflection about the x-axis	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	
Reflection about the y-axis	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$	
Reflection about the line $y = x$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	

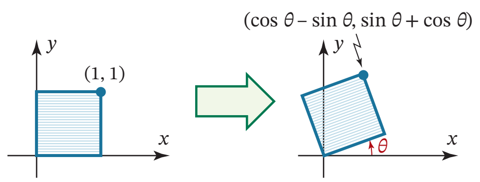
In Table 3 of Section 1.8 we obtained the standard matrices for the orthogonal projections onto the x-axis and the y-axis. **Table 2** illustrates how those projections totally flatten the unit square.

TABLE 2

Operator	Standard Matrix	Effect on the Unit Square
Orthogonal projection onto the x-axis	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	
Orthogonal projection onto the y-axis	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$	

In Table 5 of Section 1.8 we obtained the standard matrix for a rotation about the origin through a positive angle θ. **Table 3** illustrates how such a rotation transforms the unit square.

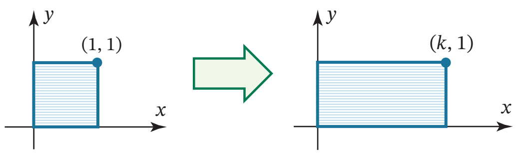
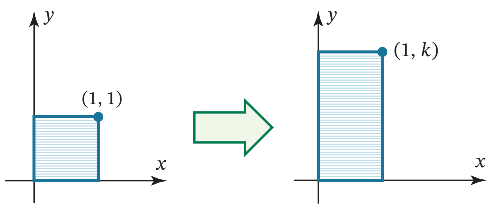
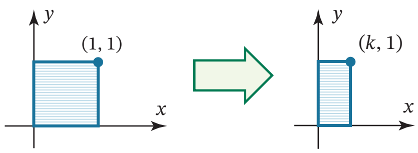
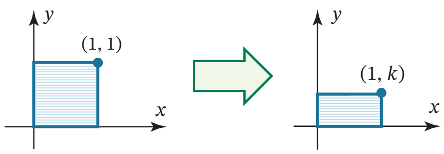
TABLE 3

Operator	Standard Matrix	Effect on the Unit Square
Rotation about the origin through a positive angle θ	$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$	

Expansions and Compressions

If the x-coordinate of each point in the plane is multiplied by a constant k where $k > 1$, the effect is to expand the unit square in the x-direction, and if the y-coordinate is multiplied by such k, then the effect is to expand the unit square in the y-direction. Operators of this type are called ***expansions***. In contrast, if $0 < k < 1$ then the effect is to compress the unit square, so such operators are called ***compressions***. **Table 4** illustrates the effect on the unit square of expansions and compressions.

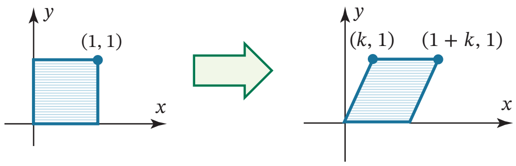
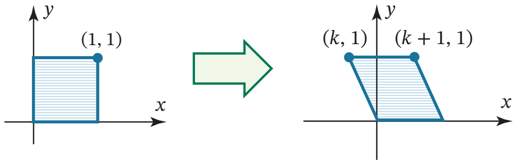
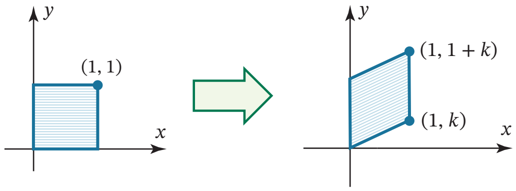
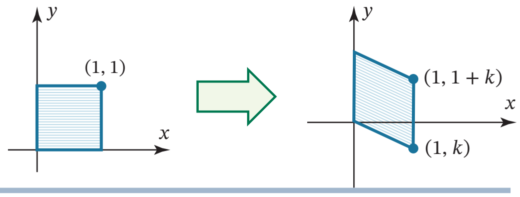
TABLE 4

Operator	Standard Matrix	Effect on the Unit Square
Expansion in the x-direction with factor k $(k > 1)$	$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$	
Expansion in the y-direction with factor k $(k > 1)$	$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$	
Compression in the x-direction with factor k $(0 < k < 1)$	$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$	
Compression in the y-direction with factor k $(0 < k < 1)$	$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$	

Shears

A matrix operator of the form $T(x, y) = (x + ky, y)$ translates a point (x, y) in the xy-plane parallel to the x-axis by an amount ky that is proportional to the y-coordinate of the point. This operator leaves the points on the x-axis fixed (since $y = 0$), but as we progress away from the x-axis, the translation distance increases. We call this operator the ***shear in the x-direction by a factor k***. Similarly, a matrix operator of the form $T(x, y) = (x, y + kx)$ is called the ***shear in the y-direction by a factor k***. **Table 5**, which illustrates the basic information about shears in R^2, shows that a shear is in the positive direction if $k > 0$ and the negative direction if $k < 0$.

TABLE 5

Operator	Standard Matrix	Effect on the Unit Square
Shear in the positive x-direction by a factor k $(k > 0)$	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$	
Shear in the negative x-direction by a factor k $(k < 0)$	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$	
Shear in the positive y-direction by a factor k $(k > 0)$	$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$	
Shear in the negative y-direction by a factor k $(k < 0)$	$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$	

EXAMPLE 3 | Transformation of the Unit Square

(*a*) Find the standard matrix for the operator on R^2 that first shears by a factor of 2 in the x-direction and then reflects the result about the line $y = x$. Sketch the image of the unit square under this operator.

(*b*) Find the standard matrix for the operator on R^2 that first reflects about $y = x$ and then shears by a factor of 2 in the x-direction. Sketch the image of the unit square under this operator.

(*c*) Confirm algebraically and visually that the shear and the reflection in parts (a) and (b) do not commute.

Solution (*a*) The standard matrix for the shear is

$$A_1 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

and for the reflection is

$$A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Thus, the standard matrix for the shear followed by the reflection is

$$A_2A_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$$

Solution (*b*) The standard matrix for the reflection followed by the shear is

$$A_1A_2 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$$

Solution (*c*) The computations in Solutions (*a*) and (*b*) show that $A_1A_2 \neq A_2A_1$, so the standard matrices, and hence the operators, do not commute. The same conclusion follows from **Figures 8.6.4** and **8.6.5** since the two operators produce different images of the unit square.

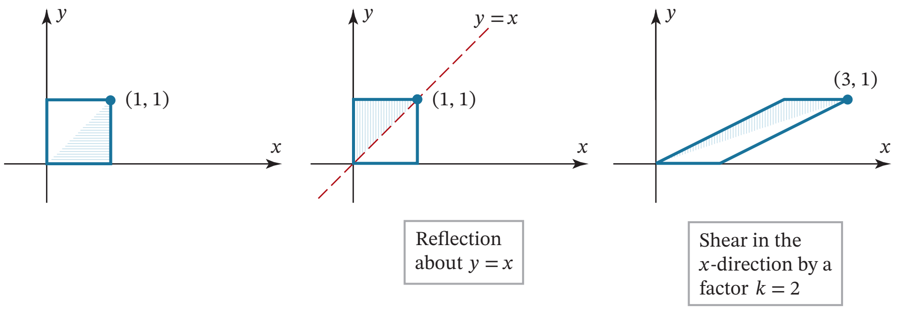


FIGURE 8.6.4

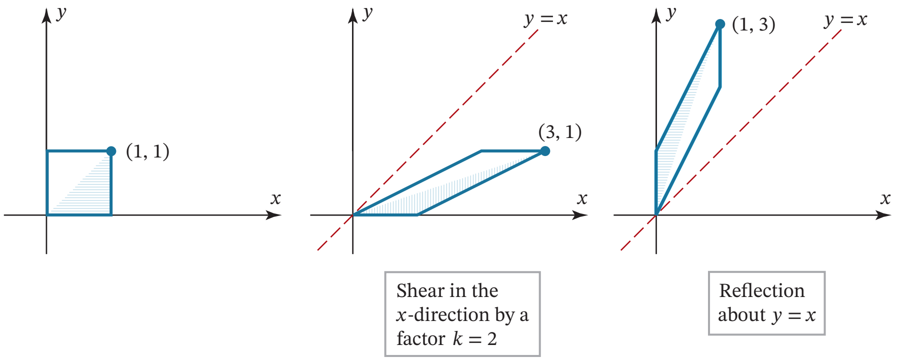


FIGURE 8.6.5

Dilations and Contractions

If k is a nonnegative scalar, then the operator $T(\mathbf{x}) = k\mathbf{x}$ on R^2 or R^3 has the effect of increasing or decreasing the length of each vector by a factor of k. If $0 \leq k < 1$ the operator is called a ***contraction*** with factor k, and if $k > 1$ it is called a ***dilation*** with factor k (**Figure 8.6.6**). If $k = 1$, then T is the identity operator. **Table 6** illustrates the effect of contractions and dilations on the unit square in R^2— contractions shrink the square and dilations enlarge it.

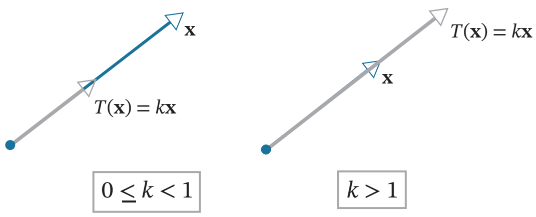
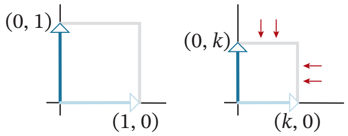
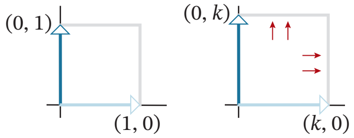


FIGURE 8.6.6

TABLE 6

Operator	Effect on the Unit Square	Standard Matrix
Contraction with factor k in R^2 $(0 \le k < 1)$		$\begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$
Dilation with factor k in R^2 $(k > 1)$		

Remark Note that the standard matrix for both contractions and dilations is a diagonal matrix all of whose diagonal entries are nonnegative and the same. A matrix operator with this property is called a ***uniform scaling***.

EXAMPLE 4 | Transformations with Diagonal Matrices

Discuss the geometric effect on the unit square of multiplication by a diagonal matrix

$$A = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}$$

in which the entries k_1 and k_2 are positive real numbers $(\neq 1)$.

Solution In the special case where $k_1 = k_2$, multiplication by A is either a contraction or dilation. More generally, the effect of multiplication by A can be seen by observing that this matrix can be factored as

$$A = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & k_2 \end{bmatrix}\begin{bmatrix} k_1 & 0 \\ 0 & 1 \end{bmatrix}$$

which shows that multiplication by A causes a compression or expansion of the unit square by a factor of k_1 in the x-direction followed by an expansion or compression of the unit square by a factor of k_2 in the y-direction.

EXAMPLE 5 | Reflection About the Origin

As illustrated in **Figure 8.6.7**, multiplication by the matrix

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

has the geometric effect of reflecting the unit square about the origin. Note, however, that the matrix equation

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

together with Table 1 shows that the same result can be obtained by first reflecting the unit square about the x-axis and then reflecting that result about the y-axis. You should be able to see this as well from **Figure 8.6.7**.

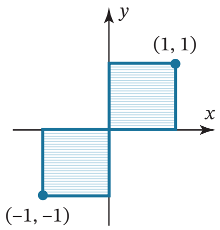


FIGURE 8.6.7

EXAMPLE 6 | Reflection About the Line $y = -x$

We leave it for you to verify that multiplication by the matrix

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

reflects the unit square about the line $y = -x$ (**Figure 8.6.8**).

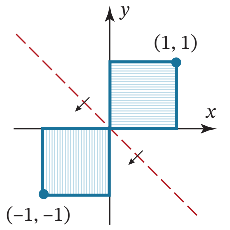


FIGURE 8.6.8

Decomposing Invertible Matrix Transformations

Our next objective is to show that all invertible matrix transformations of R^2 are expressible as compositions of compressions, expansions, reflections, and shears, so in this sense these four simple transformations are the building blocks of even the most complicated invertible matrix transformations of R^2. We begin with the following theorem.

Theorem 8.6.2

If E is an elementary matrtix, then $T_E : R^2 \to R^2$ is one of the following:

(a) A shear along a coordinate axis.

(b) A reflection about $y = x$.

(c) A compression along a coordinate axis.

(d) An expansion along a coordinate axis.

(e) A reflection about a coordinate axis.

(f) A compression or expansion along a coordinate axis followed by a reflection about a coordinate axis.

Proof Because a 2×2 elementary matrix results from performing a single elementary row operation on the 2×2 identity matrix, such a matrix must have one of the following forms (verify):

$$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$$

The first two matrices represent shears along coordinate axes, and the third represents a reflection about $y = x$. If $k > 0$, the last two matrices represent compressions or expansions along coordinate axes, depending on whether $0 \leq k < 1$ or $k > 1$. If $k < 0$, and if we express k in the form $k = -k_1$, where $k_1 > 0$, then the last two matrices can be written as

$$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -k_1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k_1 & 0 \\ 0 & 1 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -k_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & k_1 \end{bmatrix} \tag{2}$$

Since $k_1 > 0$, the product in (1) represents a compression or expansion along the x-axis followed by a reflection about the y-axis, and (2) represents a compression or expansion along the y-axis followed by a reflection about the x-axis. In the case where $k = -1$, transformations (1) and (2) are simply reflections about the y-axis and x-axis, respectively. ■

We know from Theorem 1.5.3(d) that an invertible matrix can be expressed as a product of elementary matrices, so Theorem 8.6.2 implies the following result.

Theorem 8.6.3

If $T_A : R^2 \to R^2$ is multiplication by an invertible matrix A, then the geometric effect of T_A is the same as an appropriate succession of shears, compressions, expansions, and reflections.

The next example will illustrate how Theorems 8.6.2 and 8.6.3 together with Tables 1 through 5 can be used to analyze the geometric effect of multiplication by a 2×2 invertible matrix.

EXAMPLE 7 | Decomposing a Matrix Operator

In Example 2 we illustrated the effect on the unit square of multiplication by

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

(see **Figure 8.6.3**). Express this matrix as a product of elementary matrices, and then describe the effect of multiplication by the matrix A in terms of shears, compressions, expansions, and reflections.

Solution The matrix A can be reduced to the identity matrix as follows:

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

Interchange the first and second rows. | Multiply the first row by $\frac{1}{2}$. | Add $-\frac{1}{2}$ times the second row to the first.

These three successive row operations can be performed by multiplying A on the left successively by

$$E_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad E_2 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 1 \end{bmatrix}$$

Inverting these matrices and using Formula (4) of Section 1.5 yields

$$A = \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix} = E_1^{-1}E_2^{-1}E_3^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$$

Reading from right to left we can now see that the geometric effect of multiplying by A is equivalent to successively

1. shearing by a factor of $\frac{1}{2}$ in the x-direction;
2. expanding by a factor of 2 in the x-direction;
3. reflecting about the line $y = x$.

This is illustrated in **Figure 8.6.9**, whose end result agrees with that in Example 2.

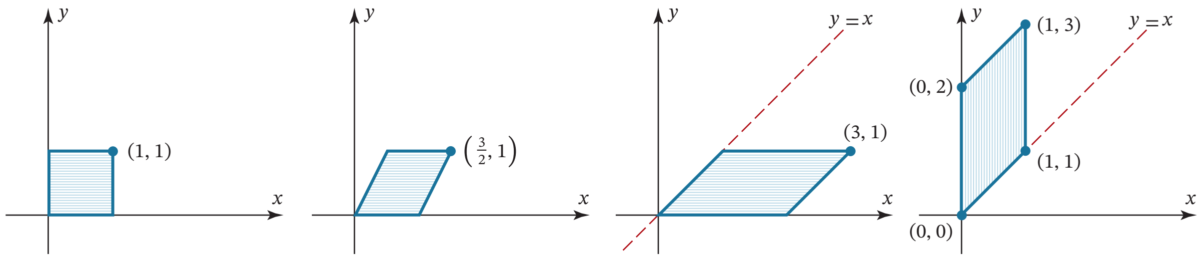
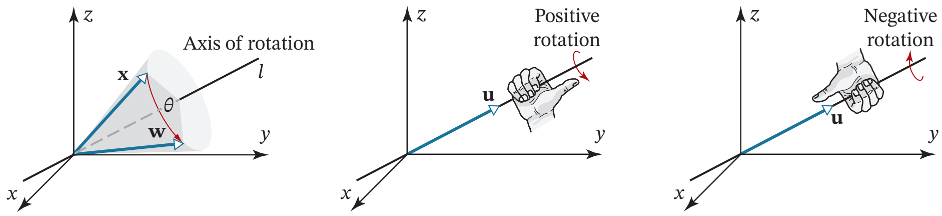


FIGURE 8.6.9

Remark Translations in R^2 are important operations in computer graphics, but they are not linear transformations (Example 8 of Section 8.1) and hence cannot be accomplished using 2×2 matrices as multipliers. A method for performing translations using 3×3 matrices is discussed in Supplementary Exercise 26.

Rotations in R^3

A rotation of vectors in R^3 is commonly described in relation to a line through the origin called the ***axis of rotation*** and a unit vector $\mathbf{u}$ along that line (**Figure 8.6.10*a***). The unit vector and what is called the ***right-hand rule*** can be used to establish a sign for the angle of rotation by cupping the fingers of your right hand so they curl in the direction of rotation and observing the direction of your thumb. If your thumb points in the direction of $\mathbf{u}$, then the angle of rotation is regarded to be ***positive*** relative to $\mathbf{u}$, and if it points in the direction opposite to $\mathbf{u}$, then it is regarded to be ***negative*** relative to $\mathbf{u}$ (**Figure 8.6.10*b***).



(*a*) Angle of rotation (*b*) Right-hand rule

FIGURE 8.6.10

For rotations about the coordinate axes in R^3, we will take the unit vectors to be $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, in which case an angle of rotation will be positive if it is counterclockwise looking toward the origin along the positive coordinate axis and will be negative if it is clockwise. **Table 7** shows the standard matrices for the ***rotation operators*** on R^3 that rotate each vector about one of the coordinate axes through an angle θ. You will find it instructive to compare these matrices to those in Table 5 of Section 1.8.

Yaw, Pitch, and Roll

In aeronautics and astronautics, the orientation of an aircraft or space shuttle relative to an xyz-coordinate system is often described in terms of angles called ***yaw***, ***pitch***, and ***roll***. If, for example, an aircraft is flying along the y-axis and the xy-plane defines the horizontal, then the aircraft's angle of rotation about the z-axis is called the ***yaw***, its angle of rotation about the x-axis is called the ***pitch***, and its angle of rotation about the y-axis is called the ***roll***. A combination of yaw, pitch, and roll can be achieved by a single rotation about some axis through the origin. This is, in fact, how a space shuttle makes attitude adjustments—it doesn't perform each rotation separately; it calculates one axis, and rotates about that axis to get the correct orientation. Such rotation maneuvers are used to align an antenna, point the nose toward a celestial object, or position a payload bay for docking.

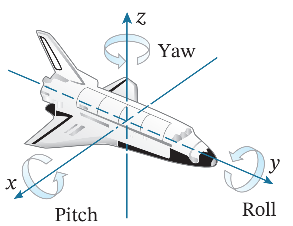
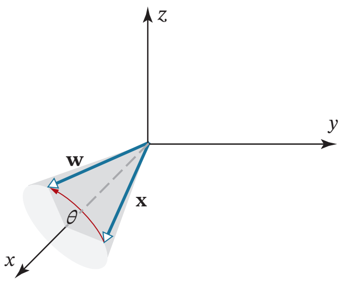
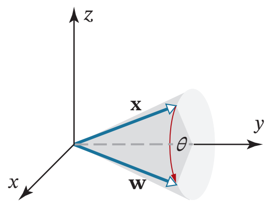
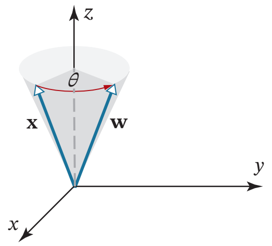


TABLE 7

Operator	Illustration	Rotation Equations	Standard Matrix
Counterclockwise rotation about the positive x-axis through an angle θ		$w_1 = x$ $w_2 = y\cos\theta - z\sin\theta$ $w_3 = y\sin\theta + z\cos\theta$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$
Counterclockwise rotation about the positive y-axis through an angle θ		$w_1 = x\cos\theta + z\sin\theta$ $w_2 = y$ $w_3 = -x\sin\theta + z\cos\theta$	$\begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$
Counterclockwise rotation about the positive z-axis through an angle θ		$w_1 = x\cos\theta - y\sin\theta$ $w_2 = x\sin\theta + y\cos\theta$ $w_3 = z$	$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$

For completeness, we note that the standard matrix for a counterclockwise rotation through an angle θ about an axis in R^3, which is determined by an arbitrary *unit vector* $\mathbf{u} = (a, b, c)$ that has its initial point at the origin, is

$$\begin{bmatrix} a^2(1-\cos\theta)+\cos\theta & ab(1-\cos\theta)-c\sin\theta & ac(1-\cos\theta)+b\sin\theta \\ ab(1-\cos\theta)+c\sin\theta & b^2(1-\cos\theta)+\cos\theta & bc(1-\cos\theta)-a\sin\theta \\ ac(1-\cos\theta)-b\sin\theta & bc(1-\cos\theta)+a\sin\theta & c^2(1-\cos\theta)+\cos\theta \end{bmatrix} \tag{3}$$

The derivation can be found in the book *Principles of Interactive Computer Graphics*, by W. M. Newman and R. F. Sproull (New York: McGraw-Hill, 1979). You may find it instructive to derive the results in Table 7 as special cases of this more general result.

Exercise Set 8.6

1. Use the method of Example 1 to find an equation for the image of the line $y = 4x$ under multiplication by the matrix

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

2. Use the method of Example 1 to find an equation for the image of the line $y = -4x + 3$ under multiplication by the matrix

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

In Exercises 3–4, find an equation for the image of the line $y = 2x$ that results from the stated transformation.

3. A shear by a factor 3 in the x-direction.

4. A compression with factor $\frac{1}{2}$ in the y-direction.

In Exercises 5–6, sketch the image of the unit square under multiplication by the given invertible matrix. As in Example 2, number the edges of the unit square and its image so it is clear how those edges correspond.

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

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

In each part of Exercises 7–8, find the standard matrix for a single operator that performs the stated succession of operations.

7. **a.** Compresses by a factor of $\frac{1}{2}$ in the x-direction, then expands by a factor of 5 in the y-direction.

b. Expands by a factor of 5 in the y-direction, then shears by a factor of 2 in the y-direction.

c. Reflects about $y = x$, then rotates through an angle of 180° about the origin.

8. **a.** Reflects about the y-axis, then expands by a factor of 5 in the x-direction, and then reflects about $y = x$.

b. Rotates through 30° about the origin, then shears by a factor of -2 in the y-direction, and then expands by a factor of 3 in the y-direction.

In each part of Exercises 9–10, determine whether the stated operators commute.

9. **a.** A reflection about the x-axis and a compression in the x-direction with factor $\frac{1}{3}$.

b. A reflection about the line $y = x$ and an expansion in the x-direction with factor 2.

10. **a.** A shear in the y-direction by a factor $\frac{1}{4}$ and a shear in the y-direction by a factor $\frac{3}{5}$.

b. A shear in the y-direction by a factor $\frac{1}{4}$ and a shear in the x-direction by a factor $\frac{3}{5}$.

In Exercises 11–14, express the matrix as a product of elementary matrices, and then describe the effect of multiplication by A in terms of shears, compressions, expansions, and reflections.

11. $A = \begin{bmatrix} 4 & 4 \\ 0 & -2 \end{bmatrix}$ **12.** $A = \begin{bmatrix} 1 & 4 \\ 2 & 9 \end{bmatrix}$

13. $A = \begin{bmatrix} 0 & -2 \\ 4 & 0 \end{bmatrix}$ **14.** $A = \begin{bmatrix} 1 & -3 \\ 4 & 6 \end{bmatrix}$

In each part of Exercises 15–16, describe, in words, the effect on the unit square of multiplication by the given diagonal matrix.

15. **a.** $A = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 1 & 0 \\ 0 & -5 \end{bmatrix}$

16. **a.** $A = \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix}$ **b.** $A = \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix}$

17. **a.** Show that multiplication by

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

maps each point in the plane onto the line $y = 2x$.

b. It follows from part (a) that the noncollinear points $(1, 0)$, $(0, 1)$, $(-1, 0)$ are mapped onto a line. Does this violate part (e) of Theorem 8.6.1?

18. Find the matrix for a shear in the x-direction that transforms the triangle with vertices $(0, 0)$, $(2, 1)$, and $(3, 0)$ into a right triangle with the right angle at the origin.

19. In accordance with part (c) of Theorem 8.6.1, show that multiplication by the invertible matrix

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

maps the parallel lines $y = 3x + 1$ and $y = 3x - 2$ into parallel lines.

20. Draw a figure that shows the image of the triangle with vertices $(0, 0)$, $(1, 0)$, and $(0.5, 1)$ under a shear by a factor of 2 in the x-direction.

21. **a.** Draw a figure that shows the image of the triangle with vertices $(0, 0)$, $(1, 0)$, and $(0.5, 1)$ under multiplication by

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

b. Find a succession of shears, compressions, expansions, and reflections that produces the same image.

22. Find the endpoints of the line segment that results when the line segment from $P(1, 2)$ to $Q(3, 4)$ is transformed by

a. a compression with factor $\frac{1}{2}$ in the y-direction.

b. a rotation of 30° about the origin.

23. Draw a figure showing the italicized letter "T" that results when the letter in the accompanying figure is sheared by a factor $\frac{1}{4}$ in the x-direction.

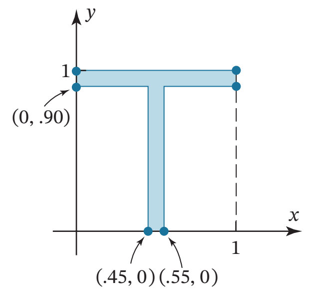


FIGURE Ex-23

24. Can an invertible matrix operator on R^2 map a square region into a triangular region? Justify your answer.

25. Find the image of the triangle with vertices $(0, 0)$, $(1, 1)$, $(2, 0)$ under multiplication by

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

Does your answer violate part (e) of Theorem 8.6.1? Explain.

26. In R^3 the ***shear in the xy-direction by a factor k*** is the matrix transformation that moves each point (x, y, z) parallel to the xy-plane to the new position $(x + kz, y + kz, z)$. (See the accompanying figure.)

a. Find the standard matrix for the shear in the xy-direction by a factor k.

b. How would you define the shear in the xz-direction by a factor k and the shear in the yz-direction by a factor k? What are the standard matrices for these matrix transformations?

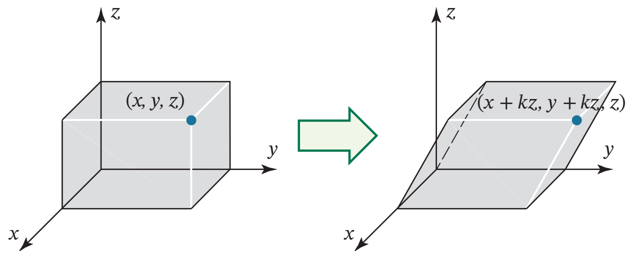


FIGURE Ex-26

In Exercises 27–28, find the standard matrix for the operator $T : R^3 \to R^3$ that performs the stated rotation.

27. a. rotates each vector 30° counterclockwise about the z-axis (looking along the positive z-axis toward the origin).

b. rotates each vector 45° clockwise about the x-axis (looking along the positive x-axis toward the origin).

28. a. rotates each vector 90° counterclockwise about the y-axis (looking along the positive y-axis toward the origin).

b. rotates each vector 90° clockwise about the positive z-axis looking toward the origin.

29. Use Formula (3) to find the standard matrix for a rotation of 180° about the axis determined by the vector $\mathbf{v} = (2, 2, 1)$. [*Note:* Formula (3) requires that the vector defining the axis of rotation have length 1.]

30. Use Formula (3) to find the standard matrix for a rotation of $\pi/2$ radians about the axis determined by $\mathbf{v} = (1, 1, 1)$. [*Note:* Formula (3) requires that the vector defining the axis of rotation have length 1.]

31. Use Formula (3) to derive the standard matrices for the rotations about the x-axis, the y-axis, and the z-axis through an angle of 90° in R^3.

32. ***Euler's Axis of Rotation Theorem*** states: *If A is an orthogonal 3×3 matrix for which $\det(A) = 1$, then multiplication by A is a rotation about a line through the origin in R^3. Moreover, if $\mathbf{u}$ is a unit vector along this line, then $A\mathbf{u} = \mathbf{u}$.*

a. Confirm that the following matrix A is orthogonal, that $\det(A) = 1$, and that there is a unit vector $\mathbf{u}$ for which $A\mathbf{u} = \mathbf{u}$.

$$A = \begin{bmatrix} \frac{2}{7} & \frac{3}{7} & \frac{6}{7} \\ \frac{3}{7} & -\frac{6}{7} & \frac{2}{7} \\ \frac{6}{7} & \frac{2}{7} & -\frac{3}{7} \end{bmatrix}$$

b. Use Formula (3) to prove that if A is a 3×3 orthogonal matrix for which $\det(A) = 1$, then the angle of rotation resulting from multiplication by A satisfies the equation $\cos\theta = \frac{1}{2}[\operatorname{tr}(A) - 1]$. Use this result to find the angle of rotation for the rotation matrix in part (a).

Working with Proofs

33. Prove part (a) of Theorem 8.6.1. [*Hint:* A line in the plane has an equation of the form $Ax + By + C = 0$, where A and B are not both zero. Use the method of Example 1 to show that the image of this line under multiplication by the invertible matrix

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

has the equation $A'x + B'y + C = 0$, where

$$A' = (dA - cB)/(ad - bc)$$

and

$$B' = (-bA + aB)/(ad - bc)$$

Then show that A' and B' are not both zero to conclude that the image is a line.

34. Use the hint in Exercise 33 to prove parts (b) and (c) of Theorem 8.6.1.

True-False Exercises

TF. In parts **(a)–(g)** determine whether the statement is true or false, and justify your answer.

a. The image of the unit square under a one-to-one matrix operator is a square.

b. A 2×2 invertible matrix operator has the geometric effect of a succession of shears, compressions, expansions, and reflections.

c. The image of a line under an invertible matrix operator is a line.

d. Every reflection operator on R^2 is its own inverse.

e. The matrix $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ represents reflection about a line.

f. The matrix $\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$ represents a shear.

g. The matrix $\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$ represents an expansion.

Working with Technology

T1. a. Find the standard matrix for the linear operator on R^3 that performs a counterclockwise rotation of 47° about the x-axis, followed by a counterclockwise rotation of 68° about the y-axis, followed by a counterclockwise rotation of 33° about the z-axis.

b. Find the image of the point $(1, 1, 1)$ under the operator in part (a).

T2. Find the standard matrix for the linear operator on R^2 that first reflects each point in the plane about the line through the origin that makes an angle of 27° with the positive x-axis and then projects the resulting point orthogonally onto the line through the origin that makes an angle of 51° with the positive x-axis.

Chapter 8 Supplementary Exercises

1. Let A be an $n \times n$ matrix, B a nonzero $n \times 1$ matrix, and $\mathbf{x}$ a vector in R^n expressed in matrix notation. Is $T(\mathbf{x}) = A\mathbf{x} + B$ a linear operator on R^n? Justify your answer.

2. Let
$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

a. Show that
$$A^2 = \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix} \quad \text{and} \quad A^3 = \begin{bmatrix} \cos 3\theta & -\sin 3\theta \\ \sin 3\theta & \cos 3\theta \end{bmatrix}$$

b. Based on your answer to part (a), make a guess at the form of the matrix A^n for any positive integer n.

c. By considering the geometric effect of multiplication by A, obtain the result in part (b) geometrically.

3. Devise a method for finding two $n \times n$ matrices that are not similar. Use your method to find two 3×3 matrices that are not similar.

4. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ be fixed vectors in R^n, and let $T: R^n \to R^m$ be the function defined by $T(\mathbf{x}) = (\mathbf{x}\cdot\mathbf{v}_1, \mathbf{x}\cdot\mathbf{v}_2, \ldots, \mathbf{x}\cdot\mathbf{v}_m)$, where $\mathbf{x}\cdot\mathbf{v}_i$ is the Euclidean inner product on R^n.

a. Show that T is a linear transformation.

b. Show that the matrix with row vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ is the standard matrix for T.

5. Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ be the standard basis for the vector space R^4, and let $T: R^4 \to R^3$ be the linear transformation for which
$$T(\mathbf{e}_1) = (1, 2, 1), \quad T(\mathbf{e}_2) = (0, 1, 0),$$
$$T(\mathbf{e}_3) = (1, 3, 0), \quad T(\mathbf{e}_4) = (1, 1, 1)$$

a. Find bases for the range and kernel of T.

b. Find the rank and nullity of T.

6. Suppose that vectors in R^3 are denoted by 1×3 matrices, and define $T: R^3 \to R^3$ by
$$T([x_1 \quad x_2 \quad x_3]) = [x_1 \quad x_2 \quad x_3] \begin{bmatrix} -1 & 2 & 4 \\ 3 & 0 & 1 \\ 2 & 2 & 5 \end{bmatrix}$$

a. Find a basis for the kernel of T.

b. Find a basis for the range of T.

7. Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ be a basis for a vector space V, and let $T: V \to V$ be the linear operator for which
$$T(\mathbf{v}_1) = \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 + 3\mathbf{v}_4$$
$$T(\mathbf{v}_2) = \mathbf{v}_1 - \mathbf{v}_2 + 2\mathbf{v}_3 + 2\mathbf{v}_4$$
$$T(\mathbf{v}_3) = 2\mathbf{v}_1 - 4\mathbf{v}_2 + 5\mathbf{v}_3 + 3\mathbf{v}_4$$
$$T(\mathbf{v}_4) = -2\mathbf{v}_1 + 6\mathbf{v}_2 - 6\mathbf{v}_3 - 2\mathbf{v}_4$$

a. Find the rank and nullity of T.

b. Determine whether T is one-to-one.

8. Let V and W be vector spaces, let T, T_1, and T_2 be linear transformations from V to W, and let k be a scalar. Define new transformations, $T_1 + T_2$ and kT, by the formulas
$$(T_1 + T_2)(\mathbf{x}) = T_1(\mathbf{x}) + T_2(\mathbf{x})$$
$$(kT)(\mathbf{x}) = k(T(\mathbf{x}))$$

a. Show that $(T_1 + T_2): V \to W$ and $kT: V \to W$ are both linear transformations.

b. Show that the set of all linear transformations from V to W with the operations in part (a) is a vector space.

9. Let A and B be similar matrices. Prove:

a. A^T and B^T are similar.

b. If A and B are invertible, then A^{-1} and B^{-1} are similar.

10. ***(Fredholm Alternative Theorem)*** Let $T: V \to V$ be a linear operator on an n-dimensional vector space. Prove that *exactly one* of the following statements holds:

i. The equation $T(\mathbf{x}) = \mathbf{b}$ has a solution for all vectors $\mathbf{b}$ in V.

ii. Nullity of $T > 0$.

11. Let $T: M_{22} \to M_{22}$ be the linear operator defined by
$$T(X) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} X + X \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$
Find the rank and nullity of T.

12. Prove: If A and B are similar matrices, and if B and C are also similar matrices, then A and C are similar matrices.

13. Let $L: M_{22} \to M_{22}$ be the linear operator that is defined by $L(M) = M^T$. Find the matrix for L with respect to the standard basis for M_{22}.

14. Let $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and $B' = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ be bases for a vector space V, and let
$$P = \begin{bmatrix} 2 & -1 & 3 \\ 1 & 1 & 4 \\ 0 & 1 & 2 \end{bmatrix}$$
be the transition matrix from B' to B.

a. Express $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ as linear combinations of $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$.

b. Express $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ as linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$.

15. Let $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ be a basis for a vector space V, and let $T: V \to V$ be a linear operator for which
$$[T]_B = \begin{bmatrix} -3 & 4 & 7 \\ 1 & 0 & -2 \\ 0 & 1 & 0 \end{bmatrix}$$
Find $[T]_{B'}$, where $B' = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is the basis for V defined by
$$\mathbf{v}_1 = \mathbf{u}_1, \quad \mathbf{v}_2 = \mathbf{u}_1 + \mathbf{u}_2, \quad \mathbf{v}_3 = \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3$$

16. Show that the matrices
$$\begin{bmatrix} 1 & 1 \\ -1 & 4 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$
are similar but that
$$\begin{bmatrix} 3 & 1 \\ -6 & -2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1 & 2 \\ 1 & 0 \end{bmatrix}$$
are not.

17. Suppose that $T: V \to V$ is a linear operator, and B is a basis for V for which

$$[T(\mathbf{x})]_B = \begin{bmatrix} x_1 - x_2 + x_3 \\ x_2 \\ x_1 - x_3 \end{bmatrix} \quad \text{if} \quad [\mathbf{x}]_B = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Find $[T]_B$.

18. Let $T: V \to V$ be a linear operator. Prove that T is one-to-one if and only if $\det(T) \neq 0$.

19. ***(Calculus required)***

a. Show that if $\mathbf{f} = f(x)$ is twice differentiable, then the function $D: C^2(-\infty, \infty) \to F(-\infty, \infty)$ defined by the formula $D(\mathbf{f}) = f''(x)$ is a linear transformation.

b. Find a basis for the kernel of D.

c. Show that the set of functions satisfying the equation $D(\mathbf{f}) = f(x)$ is a two-dimensional subspace of $C^2(-\infty, \infty)$, and find a basis for this subspace.

20. Let $T: P_2 \to R^3$ be the function defined by the formula

$$T(p(x)) = \begin{bmatrix} p(-1) \\ p(0) \\ p(1) \end{bmatrix}$$

a. Find $T(x^2 + 5x + 6)$.

b. Show that T is a linear transformation.

c. Show that T is one-to-one.

d. Find $T^{-1}(0, 3, 0)$.

e. Sketch the graph of the polynomial in part (d).

21. Let x_1, x_2, and x_3 be distinct real numbers such that

$$x_1 < x_2 < x_3$$

and let $T: P_2 \to R^3$ be the function defined by the formula

$$T(p(x)) = \begin{bmatrix} p(x_1) \\ p(x_2) \\ p(x_3) \end{bmatrix}$$

a. Show that T is a linear transformation.

b. Show that T is one-to-one.

c. Verify that if a_1, a_2, and a_3 are any real numbers, then

$$T^{-1}\left(\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}\right) = a_1P_1(x) + a_2P_2(x) + a_3P_3(x)$$

where

$$P_1(x) = \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)}$$

$$P_2(x) = \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)}$$

$$P_3(x) = \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)}$$

d. What relationship exists between the graph of the function

$$a_1P_1(x) + a_2P_2(x) + a_3P_3(x)$$

and the points (x_1, a_1), (x_2, a_2), and (x_3, a_3)?

22. ***(Calculus required)*** Let $p(x)$ and $q(x)$ be continuous functions, and let V be the subspace of $C(-\infty, \infty)$ consisting of all twice differentiable functions. Define $L: V \to V$ by

$$L(y(x)) = y''(x) + p(x)y'(x) + q(x)y(x)$$

a. Show that L is a linear transformation.

b. Consider the special case where $p(x) = 0$ and $q(x) = 1$. Show that the function

$$\phi(x) = c_1 \sin x + c_2 \cos x$$

is in the kernel of L for all real values of c_1 and c_2.

23. ***(Calculus required)*** Let $D: P_n \to P_n$ be the differentiation operator $D(\mathbf{p}) = \mathbf{p}'$. Show that the matrix for D relative to the basis $B = \{1, x, x^2, \ldots, x^n\}$ is

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & n \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

24. ***(Calculus required)*** It can be shown that for any real number c, the vectors

$$1, \quad x - c, \quad \frac{(x-c)^2}{2!}, \ldots, \quad \frac{(x-c)^n}{n!}$$

form a basis for P_n. Find the matrix for the differentiation operator of Exercise 23 with respect to this basis.

25. ***(Calculus required)*** Let $J: P_n \to P_{n+1}$ be the integration transformation defined by

$$\begin{aligned} J(\mathbf{p}) &= \int_0^x (a_0 + a_1 t + \cdots + a_n t^n)\, dt \\ &= a_0 x + \frac{a_1}{2}x^2 + \cdots + \frac{a_n}{n+1}x^{n+1} \end{aligned}$$

where $\mathbf{p} = a_0 + a_1 x + \cdots + a_n x^n$. Find the matrix for J with respect to the standard bases for P_n and P_{n+1}.

26. This exercise illustrates a method for using a matrix transformation in R^3 to translate a point (x, y) in R^2 to a point $(x + x_0,\ y + y_0)$.

a. Let

$$\mathbf{v} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{w} = \begin{bmatrix} x + x_0 \\ y + y_0 \\ 1 \end{bmatrix}$$

Find a 3×3 matrix M for which $M\mathbf{v} = \mathbf{w}$. The first top two entries in $\mathbf{w}$ are the coordinates of the translated point.

b. Use the result in part (a) to find a 3×3 matrix M that translates the point $(2, 1)$ to the point $(3, 4)$.

CHAPTER 9

Numerical Methods

CHAPTER CONTENTS

Introduction

This chapter is concerned with "numerical methods" of linear algebra, an area of study that encompasses techniques for solving large-scale linear systems and for finding numerical approximations of various kinds. It is not our objective to discuss algorithms and technical issues in fine detail since there are many excellent books on the subject. Rather, we will be concerned with introducing some of the basic ideas and exploring two important contemporary applications that rely heavily on numerical ideas—singular value decomposition and data compression. A computing utility such as MATLAB, *Mathematica*, or Maple is recommended for Sections 9.2 to 9.5.

9.1 *LU*-Decompositions

Up to now, we have focused on two methods for solving linear systems, Gaussian elimination (reduction to row echelon form) and Gauss–Jordan elimination (reduction to reduced row echelon form). While these methods are fine for the small-scale problems in this text, they are not suitable for large-scale problems in which computer roundoff error, memory usage, and speed are concerns. In this section we will discuss a method for solving a linear system of n equations in n unknowns that is based on factoring its coefficient matrix into a product of lower and upper triangular matrices. This method, called "*LU*-decomposition," is the basis for many computer algorithms in common use.

Solving Linear Systems by Factoring

Our first goal in this section is to show how to solve a linear system $A\mathbf{x} = \mathbf{b}$ of n equations in n unknowns by factoring the coefficient matrix A. We begin with some terminology.

Definition 1

A factorization of a square matrix A as

$$A = LU \tag{1}$$

where the matrix L is lower triangular and the matrix U is upper triangular, is called an ***LU-decomposition*** (or ***LU-factorization***) of A.

Before we consider the problem of obtaining an LU-decomposition, we will explain how such decompositions can be used to solve linear systems, and we will give an illustrative example.

The Method of *LU*-Decomposition

Step 1. Rewrite the system $A\mathbf{x} = \mathbf{b}$ as

$$LU\mathbf{x} = \mathbf{b} \tag{2}$$

Step 2. Make the substitution

$$\mathbf{y} = U\mathbf{x} \tag{3}$$

then rewrite (2) as $L\mathbf{y} = \mathbf{b}$ and solve this system for $\mathbf{y}$.

Step 3. Substitute $\mathbf{y}$ in (3) and solve for $\mathbf{x}$.

This procedure, which is illustrated in **Figure 9.1.1**, replaces the single linear system $A\mathbf{x} = \mathbf{b}$ by a pair of linear systems

$$\begin{aligned} U\mathbf{x} &= \mathbf{y} \\ L\mathbf{y} &= \mathbf{b} \end{aligned}$$

that must be solved in succession. However, since each of these systems has a triangular coefficient matrix, it generally turns out to involve no more computation to solve the two systems than to solve the original system directly.

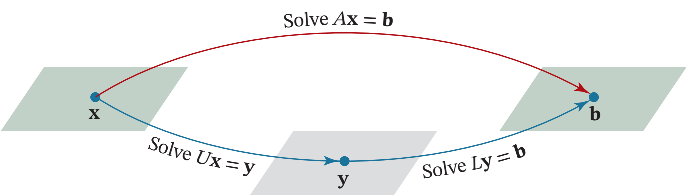


FIGURE 9.1.1

Historical Note

In 1979 an important library of machine-independent linear algebra programs called LINPACK was developed at Argonne National Laboratories. Many of the programs in that library use the decomposition methods that we will study in this section. Variations of the LINPACK routines are used in many computer programs, including MATLAB, *Mathematica*, and Maple.

EXAMPLE 1 | Solving $A\mathbf{x} = \mathbf{b}$ by *LU*-Decomposition

Later in this section we will derive the factorization

$$\underset{A}{\begin{bmatrix} 2 & 6 & 2 \\ -3 & -8 & 0 \\ 4 & 9 & 2 \end{bmatrix}} \underset{=}{=} \underset{L}{\begin{bmatrix} 2 & 0 & 0 \\ -3 & 1 & 0 \\ 4 & -3 & 7 \end{bmatrix}} \underset{U}{\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}} \tag{4}$$

Use this result to solve the linear system

$$\underset{A}{\begin{bmatrix} 2 & 6 & 2 \\ -3 & -8 & 0 \\ 4 & 9 & 2 \end{bmatrix}} \underset{\mathbf{x}}{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}} \underset{=}{=} \underset{\mathbf{b}}{\begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}}$$

From (4) we can rewrite this system as

$$\underset{L}{\begin{bmatrix} 2 & 0 & 0 \\ -3 & 1 & 0 \\ 4 & -3 & 7 \end{bmatrix}} \underset{U}{\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}} \underset{\mathbf{x}}{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}} \underset{=}{=} \underset{\mathbf{b}}{\begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}} \tag{5}$$

As specified in Step 2 above, let us define y_1, y_2, and y_3 by the equation

$$\underset{U}{\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}} \underset{\mathbf{x}}{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}} \underset{=}{=} \underset{\mathbf{y}}{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}} \tag{6}$$

which allows us to rewrite (5) as

$$\underset{L}{\begin{bmatrix} 2 & 0 & 0 \\ -3 & 1 & 0 \\ 4 & -3 & 7 \end{bmatrix}} \underset{\mathbf{y}}{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}} \underset{=}{=} \underset{\mathbf{b}}{\begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}} \tag{7}$$

or equivalently as

$$\begin{aligned} 2y_1 \phantom{{}+y_2+7y_3} &= 2 \\ -3y_1 + y_2 \phantom{{}+7y_3} &= 2 \\ 4y_1 - 3y_2 + 7y_3 &= 3 \end{aligned}$$

This system can be solved by a procedure that is similar to back substitution, except that we solve the equations from the top down instead of from the bottom up. This procedure, called ***forward substitution***, yields

$$y_1 = 1, \quad y_2 = 5, \quad y_3 = 2$$

(verify). As indicated in Step 3 above, we substitute these values into (6), which yields the linear system

$$\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix}$$

or, equivalently,

$$\begin{aligned} x_1 + 3x_2 + x_3 &= 1 \\ x_2 + 3x_3 &= 5 \\ x_3 &= 2 \end{aligned}$$

Solving this system by back substitution yields

$$x_1 = 2, \quad x_2 = -1, \quad x_3 = 2$$

(verify).

Finding *LU*-Decompositions

The preceding example illustrates that once an LU-decomposition of A is obtained, a linear system $A\mathbf{x} = \mathbf{b}$ can be solved by one forward substitution and one backward substitution. The main advantage of this method over Gaussian and Gauss–Jordan elimination is that it "decouples" A from $\mathbf{b}$ so that for solving a *sequence* of linear systems with the same coefficient matrix A, say

$$A\mathbf{x} = \mathbf{b}_1, \quad A\mathbf{x} = \mathbf{b}_2, \ldots, \quad A\mathbf{x} = \mathbf{b}_k$$

the work in factoring A need only be performed once, after which it can be reused for each system in the sequence. Such sequences occur in problems in which the matrix A remains fixed but the vector $\mathbf{b}$ varies with time.

Historical Note

Alan Mathison Turing (1912–1954)

Although the ideas were known earlier, credit for popularizing the matrix formulation of the LU-decomposition is often given to the British mathematician Alan Turing for his work on the subject in 1948. Turing, one of the great geniuses of the twentieth century, is the founder of the field of artificial intelligence. Among his many accomplishments in that field, he developed the concept of an internally programmed computer before the practical technology had reached the point where the construction of such a machine was possible. During World War II Turing was secretly recruited by the British government's Code and Cypher School at Bletchley Park to help break the Nazi Enigma codes; it was Turing's statistical approach that provided the breakthrough. In addition to being a brilliant mathematician, Turing was a world-class runner who competed successfully with Olympic-level competition. Sadly, Turing, a homosexual, was tried and convicted of "gross indecency" in 1952, in violation of the then-existing British statutes. Depressed, he committed suicide at age 41 by eating an apple laced with cyanide.

[Image: Science Source/Science Source]

Not every square matrix has an LU-decomposition. However, if it is possible to reduce a square matrix A to row echelon form by Gaussian elimination *without performing any row interchanges*, then A will have an LU-decomposition, though it may not be unique. To see why this is so, assume that A has been reduced to a row echelon form U using a sequence of row operations that does not include row interchanges. We know from Theorem 1.5.1 that these operations can be accomplished by multiplying A on the left by an appropriate sequence of elementary matrices; that is, there exist elementary matrices $E_1, E_2, \dots, E_k$ such that

$$E_k \cdots E_2 E_1 A = U \tag{8}$$

Since elementary matrices are invertible, we can solve (8) for A as

$$A = E_1^{-1} E_2^{-1} \cdots E_k^{-1} U$$

or more briefly as

$$A = LU \tag{9}$$

where

$$L = E_1^{-1} E_2^{-1} \cdots E_k^{-1} \tag{10}$$

We now have all of the ingredients to prove the following result.

Theorem 9.1.1

If A is a square matrix that can be reduced to a row echelon form U by Gaussian elimination without row interchanges, then A can be factored as $A = LU$, where L is a lower triangular matrix.

Proof Let L and U be the matrices in Formulas (10) and (8), respectively. The matrix U is upper triangular because it is a row echelon form of a square matrix (so all entries below its main diagonal are zero). To prove that L is lower triangular, it suffices to prove that each factor on the right side of (10) is lower triangular, since Theorem 1.7.1(b) will then imply that L itself is lower triangular. Since row interchanges are excluded, each E_j results either by adding a scalar multiple of one row of an identity matrix to a row below or by multiplying one row of an identity matrix by a nonzero scalar. In either case, the resulting matrix E_j is lower triangular and hence so is E_j^{-1} by Theorem 1.7.1(d). This completes the proof. ■

EXAMPLE 2 | An *LU*-Decomposition

Find an LU-decomposition of

$$A = \begin{bmatrix} 2 & 6 & 2 \\ -3 & -8 & 0 \\ 4 & 9 & 2 \end{bmatrix}$$

Solution To obtain an LU-decomposition, $A = LU$, we will reduce A to a row echelon form U using Gaussian elimination and then calculate L from (10). The steps are as follows:

	Reduction to Row Echelon Form	Row Operation	Elementary Matrix Corresponding to the Row Operation	Inverse of the Elementary Matrix
	$\begin{bmatrix} 2 & 6 & 2 \\ -3 & -8 & 0 \\ 4 & 9 & 2 \end{bmatrix}$			
Step 1.		$\frac{1}{2} \times \text{row } 1$	$E_1 = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$E_1^{-1} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
	$\begin{bmatrix} 1 & 3 & 1 \\ -3 & -8 & 0 \\ 4 & 9 & 2 \end{bmatrix}$			
Step 2.		$(3 \times \text{row } 1) + \text{row } 2$	$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$E_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
	$\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 4 & 9 & 2 \end{bmatrix}$			
Step 3.		$(-4 \times \text{row } 1) + \text{row } 3$	$E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$	$E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}$
	$\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & -3 & -2 \end{bmatrix}$			
Step 4.		$(3 \times \text{row } 2) + \text{row } 3$	$E_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix}$	$E_4^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix}$
	$\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 7 \end{bmatrix}$			
Step 5.		$\frac{1}{7} \times \text{row } 3$	$E_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{7} \end{bmatrix}$	$E_5^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix}$
	$\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} = U$			

and, from (10),

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

$$= \begin{bmatrix} 2 & 0 & 0 \\ -3 & 1 & 0 \\ 4 & -3 & 7 \end{bmatrix} \tag{11}$$

so

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

is an LU-decomposition of A.

Bookkeeping

As Example 2 shows, most of the work in constructing an LU-decomposition is expended in calculating L. However, *all* this work can be eliminated by some careful bookkeeping of the operations used to reduce A to U.

Because we are assuming that no row interchanges are required to reduce A to U, there are only two types of operations involved—multiplying a row by a nonzero constant, and adding a scalar multiple of one row to another. The first operation is used to introduce the leading 1's and the second to introduce zeros below the leading 1's.

In Example 2, a multiplier of $\frac{1}{2}$ was needed in Step 1 to introduce a leading 1 in the first row, and a multiplier of $\frac{1}{7}$ was needed in Step 5 to introduce a leading 1 in the third row. No actual multiplier was required to introduce a leading 1 in the second row because it was already a 1 at the end of Step 2, but for convenience let us say that the multiplier was 1. Comparing these multipliers with the successive diagonal entries of L, we see that these diagonal entries are precisely the reciprocals of the multipliers used to construct U:

$$L = \begin{bmatrix} 2 & 0 & 0 \\ -3 & 1 & 0 \\ 4 & -3 & 7 \end{bmatrix} \tag{12}$$

Also observe in Example 2 that to introduce zeros below the leading 1 in the first row, we used the operations

add 3 times the first row to the second
add -4 times the first row to the third

and to introduce the zero below the leading 1 in the second row, we used the operation

add 3 times the second row to the third

Now note in (11) that in each position below the main diagonal of L, the entry is the *negative* of the multiplier in the operation that introduced the zero in that position in U. This suggests the following procedure for constructing an LU-decomposition of a square matrix A, assuming that this matrix can be reduced to row echelon form without row interchanges.

Procedure for Constructing an *LU*-Decomposition

Step 1. Reduce A to a row echelon form U by Gaussian elimination without row interchanges, keeping track of the multipliers used to introduce the leading 1's and the multipliers used to introduce the zeros below the leading 1's.

Step 2. In each position along the main diagonal of L, place the reciprocal of the multiplier that introduced the leading 1 in that position in U.

Step 3. In each position below the main diagonal of L, place the negative of the multiplier used to introduce the zero in that position in U.

Step 4. Form the decomposition $A = LU$.

EXAMPLE 3 | Constructing an *LU*-Decomposition

Find an LU-decomposition of

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

Solution We will reduce A to a row echelon form U and at each step we will fill in an entry of L in accordance with the four-step procedure above.

$$A = \begin{bmatrix} 6 & -2 & 0 \\ 9 & -1 & 1 \\ 3 & 7 & 5 \end{bmatrix} \qquad \begin{bmatrix} \bullet & 0 & 0 \\ \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet \end{bmatrix}$$

• denotes an unknown entry of L.

$$\begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 9 & -1 & 1 \\ 3 & 7 & 5 \end{bmatrix} \begin{matrix} \longleftarrow \text{multiplier} = \frac{1}{6} \\ \\ \\ \end{matrix} \qquad \begin{bmatrix} 6 & 0 & 0 \\ \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 2 & 1 \\ 0 & 8 & 5 \end{bmatrix} \begin{matrix} \\ \longleftarrow \text{multiplier} = -9 \\ \longleftarrow \text{multiplier} = -3 \end{matrix} \qquad \begin{bmatrix} 6 & 0 & 0 \\ 9 & \bullet & 0 \\ 3 & \bullet & \bullet \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 8 & 5 \end{bmatrix} \begin{matrix} \\ \longleftarrow \text{multiplier} = \frac{1}{2} \\ \\ \end{matrix} \qquad \begin{bmatrix} 6 & 0 & 0 \\ 9 & 2 & 0 \\ 3 & \bullet & \bullet \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{matrix} \\ \\ \longleftarrow \text{multiplier} = -8 \end{matrix} \qquad \begin{bmatrix} 6 & 0 & 0 \\ 9 & 2 & 0 \\ 3 & 8 & \bullet \end{bmatrix}$$

$$U = \begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{matrix} \\ \\ \longleftarrow \text{multiplier} = 1 \end{matrix} \qquad L = \begin{bmatrix} 6 & 0 & 0 \\ 9 & 2 & 0 \\ 3 & 8 & 1 \end{bmatrix}$$

No actual operation is performed here since there is already a leading 1 in the third row.

Thus, we have constructed the LU-decomposition

$$A = LU = \begin{bmatrix} 6 & 0 & 0 \\ 9 & 2 & 0 \\ 3 & 8 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}$$

We leave it for you to confirm this end result by multiplying the factors.

LU-Decompositions Are Not Unique

In general, LU-decompositions are not unique. For example, if

$$A = LU = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

and L has nonzero diagonal entries (which will be true if A is invertible), then we can shift the diagonal entries from the left factor to the right factor by writing

$$\begin{aligned} A &= \begin{bmatrix} 1 & 0 & 0 \\ l_{21}/l_{11} & 1 & 0 \\ l_{31}/l_{11} & l_{32}/l_{22} & 1 \end{bmatrix} \begin{bmatrix} l_{11} & 0 & 0 \\ 0 & l_{22} & 0 \\ 0 & 0 & l_{33} \end{bmatrix} \begin{bmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ l_{21}/l_{11} & 1 & 0 \\ l_{31}/l_{11} & l_{32}/l_{22} & 1 \end{bmatrix} \begin{bmatrix} l_{11} & l_{11}u_{12} & l_{11}u_{13} \\ 0 & l_{22} & l_{22}u_{23} \\ 0 & 0 & l_{33} \end{bmatrix} \end{aligned}$$

which is another LU-decomposition of A.

LDU-Decompositions

The method we have given for computing LU-decompositions may result in an "asymmetry" in that the matrix U has 1's on the main diagonal but L need not. However, if it is preferred to have 1's on the main diagonal of both the lower triangular factor and the upper triangular factor, then we can "shift" the diagonal entries of L to a diagonal matrix D and write L as

$$L = L'D$$

where L' is a lower triangular matrix with 1's on the main diagonal. For example, a general 3×3 lower triangular matrix with nonzero entries on the main diagonal can be factored as

$$\underset{L}{\begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}} = \underset{L'}{\begin{bmatrix} 1 & 0 & 0 \\ a_{21}/a_{11} & 1 & 0 \\ a_{31}/a_{11} & a_{32}/a_{22} & 1 \end{bmatrix}} \underset{D}{\begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix}}$$

Note that the columns of L' are obtained by dividing each entry in the corresponding column of L by the diagonal entry in the column. Thus, for example, we can rewrite (4) as

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

$$= \begin{bmatrix} 1 & 0 & 0 \\ -\frac{3}{2} & 1 & 0 \\ 2 & -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix} \begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \tag{13}$$

If desired, the diagonal matrix and the upper triangular matrix in (13) can be multiplied to produce an LU-decomposition in which the 1's are on the main diagonal of L rather than U.

One can prove that if A is an invertible matrix that can be reduced to row echelon form without row interchanges, then A can be factored *uniquely* as

$$A = LDU$$

where L is a lower triangular matrix with 1's on the main diagonal, D is a diagonal matrix, and U is an upper triangular matrix with 1's on the main diagonal. This is called the ***LDU*-decomposition** (or ***LDU*-factorization**) of A.

PLU-Decompositions

Many computer algorithms for solving linear systems perform row interchanges to reduce roundoff error, in which case the existence of an LU-decomposition is not guaranteed. However, it is possible to work around this problem by "preprocessing" the coefficient matrix A so that the row interchanges are performed *prior* to computing the LU-decomposition itself. The idea is to create a matrix Q (called a ***permutation matrix***) by multiplying, in sequence, those elementary matrices that produce the row interchanges and then execute them by computing the product QA. This product can then be reduced to row echelon form *without* row interchanges, so it is assured to have an LU-decomposition

$$QA = LU \tag{14}$$

Because the matrix Q is invertible (being a product of elementary matrices), the systems $A\mathbf{x} = \mathbf{b}$ and $QA\mathbf{x} = Q\mathbf{b}$ will have the same solutions. But it follows from (14) that the latter system can be rewritten as $LU\mathbf{x} = Q\mathbf{b}$ and hence can be solved using LU-decomposition.

It is common to see Equation (14) expressed as

$$A = PLU \tag{15}$$

in which $P = Q^{-1}$. This is called a ***PLU*-decomposition** or (***PLU*-factorization**) of A.

Exercise Set 9.1

1. Use the method of Example 1 and the LU-decomposition

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

to solve the system

$$\begin{aligned} 3x_1 - 6x_2 &= 0 \\ -2x_1 + 5x_2 &= 1 \end{aligned}$$

2. Use the method of Example 1 and the LU-decomposition

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

to solve the system

$$\begin{aligned} 3x_1 - 6x_2 - 3x_3 &= -3 \\ 2x_1 \qquad + 6x_3 &= -22 \\ -4x_1 + 7x_2 + 4x_3 &= 3 \end{aligned}$$

In Exercises 3–6, find an LU-decomposition of the coefficient matrix, and then use the method of Example 1 to solve the system.

3. $\begin{bmatrix} 2 & 8 \\ -1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$

4. $\begin{bmatrix} -5 & -10 \\ 6 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -10 \\ 19 \end{bmatrix}$

5. $\begin{bmatrix} 2 & -2 & -2 \\ 0 & -2 & 2 \\ -1 & 5 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -4 \\ -2 \\ 6 \end{bmatrix}$

6. $\begin{bmatrix} -3 & 12 & -6 \\ 1 & -2 & 2 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -33 \\ 7 \\ -1 \end{bmatrix}$

In Exercises 7–8, an LU-decomposition of a matrix A is given.

a. *Compute L^{-1} and U^{-1}.*

b. *Use the result in part (a) to find the inverse of A.*

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

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

8. The LU-decomposition obtained in Example 2.

9. Let

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

a. Find an LU-decomposition of A.

b. Express A in the form $A = L_1DU_1$, where L_1 is lower triangular with 1's along the main diagonal, U_1 is upper triangular, and D is a diagonal matrix.

c. Express A in the form $A = L_2U_2$, where L_2 is lower triangular with 1's along the main diagonal and U_2 is upper triangular.

10. **a.** Show that the matrix

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

has no LU-decomposition.

b. Find a PLU-decomposition of this matrix.

In Exercises 11–12, use the given PLU-decomposition of A to solve the linear system $A\mathbf{x} = \mathbf{b}$ by rewriting it as $P^{-1}A\mathbf{x} = P^{-1}\mathbf{b}$ and solving this system by LU-decomposition.

11. $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}$; $A = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 2 & 2 \\ 3 & 1 & 3 \end{bmatrix}$;

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & -5 & 17 \end{bmatrix}\begin{bmatrix} 1 & 2 & 2 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix} = PLU$$

12. $\mathbf{b} = \begin{bmatrix} 3 \\ 0 \\ 6 \end{bmatrix}$; $A = \begin{bmatrix} 4 & 1 & 2 \\ 0 & 2 & 1 \\ 8 & 1 & 8 \end{bmatrix}$;

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 4 & 1 & 2 \\ 0 & -1 & 4 \\ 0 & 0 & 9 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{4} & \frac{1}{2} \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{bmatrix} = PLU$$

In Exercises 13–14, find the LDU-decomposition of A.

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

14. $A = \begin{bmatrix} 3 & -12 & 6 \\ 0 & 2 & 0 \\ 6 & -28 & 13 \end{bmatrix}$

In Exercises 15–16, find a PLU-decomposition of A, and use it to solve the linear system $A\mathbf{x} = \mathbf{b}$ by the method of Exercises 11 and 12.

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

16. $A = \begin{bmatrix} 0 & 3 & -2 \\ 1 & 1 & 4 \\ 2 & 2 & 5 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 7 \\ 5 \\ -2 \end{bmatrix}$

17. Let $A\mathbf{x} = \mathbf{b}$ be a linear system of n equations in n unknowns, and assume that A is an invertible matrix that can be reduced to row echelon form without row interchanges. How many additions and multiplications are required to solve the system by the method of Example 1?

Working with Proofs

18. Let

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

a. Prove: If $a \neq 0$, then the matrix A has a unique LU-decomposition with 1's along the main diagonal of L.

b. Find the LU-decomposition described in part (a).

19. Prove: If A is any $n \times n$ matrix, then A can be factored as $A = PLU$, where L is lower triangular, U is upper triangular, and P can be obtained by interchanging the rows of I_n appropriately. [*Hint:* Let U be a row echelon form of A, and let all row interchanges required in the reduction of A to U be performed first.]

True-False Exercises

TF. In parts **(a)–(e)** determine whether the statement is true or false, and justify your answer.

a. Every square matrix has an LU-decomposition.

b. If a square matrix A is row equivalent to an upper triangular matrix U, then A has an LU-decomposition.

c. If $L_1, L_2, \ldots, L_k$ are $n \times n$ lower triangular matrices, then the product $L_1 L_2 \cdots L_k$ is lower triangular.

d. If an invertible matrix A has an LU-decomposition, then A has a unique LDU-decomposition.

e. Every square matrix has a PLU-decomposition.

Working with Technology

T1. Technology utilities vary in how they handle LU-decompositions. For example, many utilities perform row interchanges to reduce roundoff error and hence produce PLU-decompositions, even when asked for LU-decompositions. See what happens when you use your utility to find an LU-decomposition of the matrix A in Example 2.

T2. The accompanying figure shows a metal plate whose edges are held at the temperatures shown. It follows from thermodynamic principles that the temperature at each of the six interior nodes will eventually stabilize at a value that is approximately the average of the temperatures at the four neighboring nodes. These are called the ***steady-state temperatures*** at the nodes. Thus, for example, if we denote the steady-state temperatures at the interior nodes in the accompanying figure as T_1, T_2, T_3, T_4, T_5, and T_6, then at the node labeled T_1 that temperature will be $T_1 = \frac{1}{4}(0 + 5 + T_2 + T_3)$ or, equivalently,

$$4T_1 - T_2 - T_3 = 5$$

Find a linear system whose solution gives the steady-state temperatures at the nodes, and use your technology utility to solve that system by LU-decomposition.

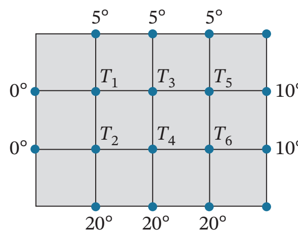


FIGURE Ex-T2

9.2 The Power Method

The eigenvalues of a square matrix can, in theory, be found by solving the characteristic equation. However, this procedure has so many computational difficulties that it is almost never used in applications. In this section we will discuss an algorithm that can be used to approximate the eigenvalue with greatest absolute value and a corresponding eigenvector. This particular eigenvalue and its corresponding eigenvectors are important because they arise naturally in many iterative processes. The methods we will study in this section have recently been used to create Internet search engines such as Google.

The Power Method

There are many applications in which some vector $\mathbf{x}_0$ in R^n is multiplied repeatedly by an $n \times n$ matrix A to produce a sequence

$$\mathbf{x}_0, \quad A\mathbf{x}_0, \quad A^2\mathbf{x}_0, \ldots, \quad A^k\mathbf{x}_0, \ldots$$

We call a sequence of this form a ***power sequence generated by A***. In this section we will be concerned with the convergence of power sequences and how such sequences can be used to approximate eigenvalues and eigenvectors. For this purpose, we make the following definition.

Definition 1

If the *distinct* eigenvalues of a matrix A are $\lambda_1, \lambda_2, \ldots, \lambda_k$, and if $|\lambda_1|$ is larger than $|\lambda_2|, \ldots, |\lambda_k|$, then λ_1 is called a ***dominant eigenvalue*** of A. Any eigenvector corresponding to a dominant eigenvalue is called a ***dominant eigenvector*** of A.

EXAMPLE 1 | Dominant Eigenvalues

Some matrices have dominant eigenvalues and some do not. For example, if the distinct eigenvalues of a matrix are

$$\lambda_1 = -4, \quad \lambda_2 = -2, \quad \lambda_3 = 1, \quad \lambda_4 = 3$$

then $\lambda_1 = -4$ is dominant since $|\lambda_1| = 4$ is greater than the absolute values of all the other eigenvalues; but if the distinct eigenvalues of a matrix are

$$\lambda_1 = 7, \quad \lambda_2 = -7, \quad \lambda_3 = -2, \quad \lambda_4 = 5$$

then $|\lambda_1| = |\lambda_2| = 7$, so there is no single eigenvalue whose absolute value is greater than the absolute value of all the other eigenvalues.

The most important theorems about convergence of power sequences apply to $n \times n$ matrices with n linearly independent eigenvectors (symmetric matrices, for example), so we will limit our discussion to this case in this section.

Theorem 9.2.1

Let A be a symmetric $n \times n$ matrix that has a positive* dominant eigenvalue λ. If $\mathbf{x}_0$ is a unit vector in R^n that is not orthogonal to the eigenspace corresponding to λ, then the normalized power sequence

$$\mathbf{x}_0, \quad \mathbf{x}_1 = \frac{A\mathbf{x}_0}{\|A\mathbf{x}_0\|}, \quad \mathbf{x}_2 = \frac{A\mathbf{x}_1}{\|A\mathbf{x}_1\|}, \ldots, \quad \mathbf{x}_k = \frac{A\mathbf{x}_{k-1}}{\|A\mathbf{x}_{k-1}\|}, \ldots \tag{1}$$

converges to a unit dominant eigenvector, and the sequence

$$A\mathbf{x}_0 \cdot \mathbf{x}_0, \quad A\mathbf{x}_1 \cdot \mathbf{x}_1, \ldots, \quad A\mathbf{x}_k \cdot \mathbf{x}_k, \ldots \tag{2}$$

converges to the dominant eigenvalue λ.

Remark In the exercises we will ask you to show that (1) can also be expressed as

$$\mathbf{x}_0, \quad \mathbf{x}_1 = \frac{A\mathbf{x}_0}{\|A\mathbf{x}_0\|}, \quad \mathbf{x}_2 = \frac{A^2\mathbf{x}_0}{\|A^2\mathbf{x}_0\|}, \ldots, \quad \mathbf{x}_k = \frac{A^k\mathbf{x}_0}{\|A^k\mathbf{x}_0\|}, \ldots \tag{3}$$

This form of the power sequence expresses each iterate in terms of the starting vector $\mathbf{x}_0$, rather than in terms of its predecessor.

We will not prove Theorem 9.2.1, but we can make it plausible geometrically in the 2×2 case where A is a symmetric matrix with distinct positive eigenvalues, λ_1 and λ_2, one of which is dominant. To be specific, assume that λ_1 is dominant and

$$\lambda_1 > \lambda_2 > 0$$

*If the dominant eigenvalue is not positive, sequence (2) will still converge to the dominant eigenvalue, but sequence (1) may not converge to a *specific* dominant eigenvector because of *alternation* (see Exercise 11). Nevertheless, each term of (1) will closely approximate *some* dominant eigenvector for sufficiently large values of k.

Since we are assuming that A is symmetric and has distinct eigenvalues, it follows from Theorem 7.2.2 that the eigenspaces corresponding to λ_1 and λ_2 are perpendicular lines through the origin. Thus, the assumption that $\mathbf{x}_0$ is a unit vector that is not orthogonal to the eigenspace corresponding to λ_1 implies that $\mathbf{x}_0$ does not lie in the eigenspace corresponding to λ_2. To see the geometric effect of multiplying $\mathbf{x}_0$ by A, it will be useful to split $\mathbf{x}_0$ into the sum

$$\mathbf{x}_0 = \mathbf{v}_0 + \mathbf{w}_0 \tag{4}$$

where $\mathbf{v}_0$ and $\mathbf{w}_0$ are the orthogonal projections of $\mathbf{x}_0$ on the eigenspaces of λ_1 and λ_2, respectively (**Figure 9.2.1*a***).

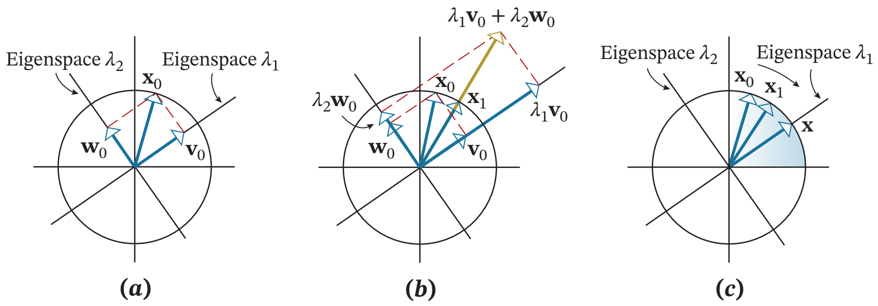


FIGURE 9.2.1

This enables us to express $A\mathbf{x}_0$ as

$$A\mathbf{x}_0 = A\mathbf{v}_0 + A\mathbf{w}_0 = \lambda_1\mathbf{v}_0 + \lambda_2\mathbf{w}_0 \tag{5}$$

which tells us that multiplying $\mathbf{x}_0$ by A "scales" the terms $\mathbf{v}_0$ and $\mathbf{w}_0$ in (4) by λ_1 and λ_2, respectively. However, λ_1 is larger than λ_2, so the scaling is greater in the direction of $\mathbf{v}_0$ than in the direction of $\mathbf{w}_0$. Thus, multiplying $\mathbf{x}_0$ by A "pulls" $\mathbf{x}_0$ toward the eigenspace of λ_1, and normalizing produces a vector $\mathbf{x}_1 = A\mathbf{x}_0 / \|A\mathbf{x}_0\|$, which is on the unit circle and is closer to the eigenspace of λ_1 than $\mathbf{x}_0$ (**Figure 9.2.1*b***). Similarly, multiplying $\mathbf{x}_1$ by A and normalizing produces a unit vector $\mathbf{x}_2$ that is closer to the eigenspace of λ_1 than $\mathbf{x}_1$. Thus, it seems reasonable that by repeatedly multiplying by A and normalizing we will produce a sequence of vectors $\mathbf{x}_k$ that lie on the unit circle and converge to a unit vector $\mathbf{x}$ in the eigenspace of λ_1 (**Figure 9.2.1*c***). Moreover, if $\mathbf{x}_k$ converges to $\mathbf{x}$, then it also seems reasonable that $A\mathbf{x}_k \cdot \mathbf{x}_k$ will converge to

$$A\mathbf{x} \cdot \mathbf{x} = \lambda_1 \mathbf{x} \cdot \mathbf{x} = \lambda_1 \|\mathbf{x}\|^2 = \lambda_1$$

which is the dominant eigenvalue of A.

The Power Method with Euclidean Scaling

Theorem 9.2.1 provides us with an algorithm for approximating the dominant eigenvalue and a corresponding unit eigenvector of a symmetric matrix A, provided the dominant eigenvalue is positive. This algorithm, called the ***power method with Euclidean scaling***, is as follows:

The Power Method with Euclidean Scaling

Step 0. Choose an arbitrary nonzero vector and normalize it, if need be, to obtain a unit vector $\mathbf{x}_0$.

Step 1. Compute $A\mathbf{x}_0$ and normalize it to obtain the first approximation $\mathbf{x}_1$ to a dominant unit eigenvector. Compute $A\mathbf{x}_1 \cdot \mathbf{x}_1$ to obtain the first approximation to the dominant eigenvalue.

Step 2. Compute $A\mathbf{x}_1$ and normalize it to obtain the second approximation $\mathbf{x}_2$ to a dominant unit eigenvector. Compute $A\mathbf{x}_2 \cdot \mathbf{x}_2$ to obtain the second approximation to the dominant eigenvalue.

Step 3. Compute $A\mathbf{x}_2$ and normalize it to obtain the third approximation $\mathbf{x}_3$ to a dominant unit eigenvector. Compute $A\mathbf{x}_3 \cdot \mathbf{x}_3$ to obtain the third approximation to the dominant eigenvalue.

Continuing in this way will usually generate a sequence of better and better approximations to the dominant eigenvalue and a corresponding unit eigenvector.*

EXAMPLE 2 | The Power Method with Euclidean Scaling

Apply the power method with Euclidean scaling to

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \quad \text{with} \quad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Stop at $\mathbf{x}_5$ and compare the resulting approximations to the exact values of the dominant eigenvalue and eigenvector.

Solution We will leave it for you to show that the eigenvalues of A are $\lambda = 1$ and $\lambda = 5$ and that the eigenspace corresponding to the dominant eigenvalue $\lambda = 5$ is the line represented by the parametric equations $x_1 = t, x_2 = t$, which we can write in vector form as

$$\mathbf{x} = t\begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{6}$$

Setting $t = 1/\sqrt{2}$ yields the normalized dominant eigenvector

$$\mathbf{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \approx \begin{bmatrix} 0.707106781187\ldots \\ 0.707106781187\ldots \end{bmatrix} \tag{7}$$

Now let us see what happens when we use the power method, starting with the unit vector $\mathbf{x}_0$.

$$A\mathbf{x}_0 = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad \mathbf{x}_1 = \frac{A\mathbf{x}_0}{\|A\mathbf{x}_0\|} = \frac{1}{\sqrt{13}}\begin{bmatrix} 3 \\ 2 \end{bmatrix} \approx \frac{1}{3.60555}\begin{bmatrix} 3 \\ 2 \end{bmatrix} \approx \begin{bmatrix} 0.83205 \\ 0.55470 \end{bmatrix}$$

$$A\mathbf{x}_1 \approx \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 0.83205 \\ 0.55470 \end{bmatrix} \approx \begin{bmatrix} 3.60555 \\ 3.32820 \end{bmatrix} \qquad \mathbf{x}_2 = \frac{A\mathbf{x}_1}{\|A\mathbf{x}_1\|} \approx \frac{1}{4.90682}\begin{bmatrix} 3.60555 \\ 3.32820 \end{bmatrix} \approx \begin{bmatrix} 0.73480 \\ 0.67828 \end{bmatrix}$$

$$A\mathbf{x}_2 \approx \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 0.73480 \\ 0.67828 \end{bmatrix} \approx \begin{bmatrix} 3.56097 \\ 3.50445 \end{bmatrix} \qquad \mathbf{x}_3 = \frac{A\mathbf{x}_2}{\|A\mathbf{x}_2\|} \approx \frac{1}{4.99616}\begin{bmatrix} 3.56097 \\ 3.50445 \end{bmatrix} \approx \begin{bmatrix} 0.71274 \\ 0.70143 \end{bmatrix}$$

$$A\mathbf{x}_3 \approx \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 0.71274 \\ 0.70143 \end{bmatrix} \approx \begin{bmatrix} 3.54108 \\ 3.52976 \end{bmatrix} \qquad \mathbf{x}_4 = \frac{A\mathbf{x}_3}{\|A\mathbf{x}_3\|} \approx \frac{1}{4.99985}\begin{bmatrix} 3.54108 \\ 3.52976 \end{bmatrix} \approx \begin{bmatrix} 0.70824 \\ 0.70597 \end{bmatrix}$$

$$A\mathbf{x}_4 \approx \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 0.70824 \\ 0.70597 \end{bmatrix} \approx \begin{bmatrix} 3.53666 \\ 3.53440 \end{bmatrix} \qquad \mathbf{x}_5 = \frac{A\mathbf{x}_4}{\|A\mathbf{x}_4\|} \approx \frac{1}{4.99999}\begin{bmatrix} 3.53666 \\ 3.53440 \end{bmatrix} \approx \begin{bmatrix} 0.70733 \\ 0.70688 \end{bmatrix}$$

*If the vector $\mathbf{x}_0$ happens to be orthogonal to the eigenspace of the dominant eigenvalue, then the hypotheses of Theorem 9.2.1 will be violated and the method may fail. However, the reality is that computer roundoff errors usually perturb $\mathbf{x}_0$ enough to destroy any orthogonality and make the algorithm work. This is one instance in which errors help to obtain correct results!

$$\lambda^{(1)} = (A\mathbf{x}_1)\cdot\mathbf{x}_1 = (A\mathbf{x}_1)^T\mathbf{x}_1 \approx \begin{bmatrix}3.60555 & 3.32820\end{bmatrix}\begin{bmatrix}0.83205\\0.55470\end{bmatrix} \approx 4.84615$$

$$\lambda^{(2)} = (A\mathbf{x}_2)\cdot\mathbf{x}_2 = (A\mathbf{x}_2)^T\mathbf{x}_2 \approx \begin{bmatrix}3.56097 & 3.50445\end{bmatrix}\begin{bmatrix}0.73480\\0.67828\end{bmatrix} \approx 4.99361$$

$$\lambda^{(3)} = (A\mathbf{x}_3)\cdot\mathbf{x}_3 = (A\mathbf{x}_3)^T\mathbf{x}_3 \approx \begin{bmatrix}3.54108 & 3.52976\end{bmatrix}\begin{bmatrix}0.71274\\0.70143\end{bmatrix} \approx 4.99974$$

$$\lambda^{(4)} = (A\mathbf{x}_4)\cdot\mathbf{x}_4 = (A\mathbf{x}_4)^T\mathbf{x}_4 \approx \begin{bmatrix}3.53666 & 3.53440\end{bmatrix}\begin{bmatrix}0.70824\\0.70597\end{bmatrix} \approx 4.99999$$

$$\lambda^{(5)} = (A\mathbf{x}_5)\cdot\mathbf{x}_5 = (A\mathbf{x}_5)^T\mathbf{x}_5 \approx \begin{bmatrix}3.53576 & 3.53531\end{bmatrix}\begin{bmatrix}0.70733\\0.70688\end{bmatrix} \approx 5.00000$$

Thus, $\lambda^{(5)}$ approximates the dominant eigenvalue to five decimal place accuracy and $\mathbf{x}_5$ approximates the dominant eigenvector in (7) to three decimal place accuracy.

It is accidental that $\lambda^{(5)}$ (the fifth approximation) produced five decimal place accuracy. In general, n iterations need not produce n decimal place accuracy.

The Power Method with Maximum Entry Scaling

There is a variation of the power method in which the iterates, rather than being normalized at each stage, are scaled to make the maximum entry 1. To describe this method, it will be convenient to denote the maximum *absolute value* of the entries in a vector $\mathbf{x}$ by $\max(\mathbf{x})$. Thus, for example, if

$$\mathbf{x} = \begin{bmatrix}5\\3\\-7\\2\end{bmatrix}$$

then $\max(\mathbf{x}) = 7$. We will need the following variation of Theorem 9.2.1.

Theorem 9.2.2

Let A be a symmetric $n \times n$ matrix that has a positive dominant* eigenvalue λ. If $\mathbf{x}_0$ is a nonzero vector in R^n that is not orthogonal to the eigenspace corresponding to λ, then the sequence

$$\mathbf{x}_0, \quad \mathbf{x}_1 = \frac{A\mathbf{x}_0}{\max(A\mathbf{x}_0)}, \quad \mathbf{x}_2 = \frac{A\mathbf{x}_1}{\max(A\mathbf{x}_1)}, \ldots, \quad \mathbf{x}_k = \frac{A\mathbf{x}_{k-1}}{\max(A\mathbf{x}_{k-1})}, \ldots \tag{8}$$

converges to an eigenvector corresponding to λ, and the sequence

$$\frac{A\mathbf{x}_0\cdot\mathbf{x}_0}{\mathbf{x}_0\cdot\mathbf{x}_0}, \quad \frac{A\mathbf{x}_1\cdot\mathbf{x}_1}{\mathbf{x}_1\cdot\mathbf{x}_1}, \ldots, \quad \frac{A\mathbf{x}_k\cdot\mathbf{x}_k}{\mathbf{x}_k\cdot\mathbf{x}_k}, \ldots \tag{9}$$

converges to λ.

Remark In the exercises we will ask you to show that (8) can be written in the alternative form

$$\mathbf{x}_0, \quad \mathbf{x}_1 = \frac{A\mathbf{x}_0}{\max(A\mathbf{x}_0)}, \quad \mathbf{x}_2 = \frac{A^2\mathbf{x}_0}{\max(A^2\mathbf{x}_0)}, \ldots, \quad \mathbf{x}_k = \frac{A^k\mathbf{x}_0}{\max(A^k\mathbf{x}_0)}, \ldots \tag{10}$$

which expresses the iterates in terms of the initial vector $\mathbf{x}_0$.

*As in Theorem 9.2.1, if the dominant eigenvalue is not positive, sequence (9) will still converge to the dominant eigenvalue, but sequence (8) may not converge to a *specific* dominant eigenvector. Nevertheless, each term of (8) will closely approximate *some* dominant eigenvector for sufficiently large values of k (see Exercise 11).

We will omit the proof of this theorem, but if we accept that (8) converges to an eigenvector of A, then it is not hard to see why (9) converges to the dominant eigenvalue. To see this, note that each term in (9) is of the form

$$\frac{A\mathbf{x}\cdot\mathbf{x}}{\mathbf{x}\cdot\mathbf{x}} \tag{11}$$

which is called a ***Rayleigh quotient*** of A. In the case where λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector, the Rayleigh quotient is

$$\frac{A\mathbf{x}\cdot\mathbf{x}}{\mathbf{x}\cdot\mathbf{x}} = \frac{\lambda\mathbf{x}\cdot\mathbf{x}}{\mathbf{x}\cdot\mathbf{x}} = \frac{\lambda(\mathbf{x}\cdot\mathbf{x})}{\mathbf{x}\cdot\mathbf{x}} = \lambda$$

Thus, if $\mathbf{x}_k$ converges to a dominant eigenvector $\mathbf{x}$, then it seems reasonable that

$$\frac{A\mathbf{x}_k\cdot\mathbf{x}_k}{\mathbf{x}_k\cdot\mathbf{x}_k} \quad \text{converges to} \quad \frac{A\mathbf{x}\cdot\mathbf{x}}{\mathbf{x}\cdot\mathbf{x}} = \lambda$$

which is the dominant eigenvalue.

Theorem 9.2.2 produces the following algorithm, which is called the ***power method with maximum entry scaling***.

The Power Method with Maximum Entry Scaling

Step 0. Choose an arbitrary nonzero vector $\mathbf{x}_0$.

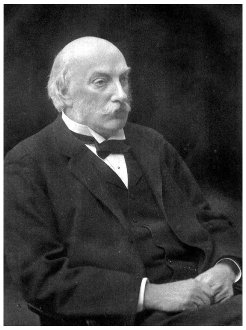
Step 1. Compute $A\mathbf{x}_0$ and multiply it by the factor $1/\max(A\mathbf{x}_0)$ to obtain the first approximation $\mathbf{x}_1$ to a dominant eigenvector. Compute the Rayleigh quotient of $\mathbf{x}_1$ to obtain the first approximation to the dominant eigenvalue.

Step 2. Compute $A\mathbf{x}_1$ and scale it by the factor $1/\max(A\mathbf{x}_1)$ to obtain the second approximation $\mathbf{x}_2$ to a dominant eigenvector. Compute the Rayleigh quotient of $\mathbf{x}_2$ to obtain the second approximation to the dominant eigenvalue.

Step 3. Compute $A\mathbf{x}_2$ and scale it by the factor $1/\max(A\mathbf{x}_2)$ to obtain the third approximation $\mathbf{x}_3$ to a dominant eigenvector. Compute the Rayleigh quotient of $\mathbf{x}_3$ to obtain the third approximation to the dominant eigenvalue.

Continuing in this way will generate a sequence of better and better approximations to the dominant eigenvalue and a corresponding eigenvector.

Historical Note



John William Strutt Rayleigh (1842–1919)

The British mathematical physicist John Rayleigh won the Nobel prize in physics in 1904 for his discovery of the inert gas argon. Rayleigh also made fundamental discoveries in acoustics and optics, and his work in wave phenomena enabled him to give the first accurate explanation of why the sky is blue.

[Image: The Granger Collection, New York]

EXAMPLE 3 | Example 2 Revisited Using Maximum Entry Scaling

Apply the power method with maximum entry scaling to

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \quad \text{with} \quad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Stop at $\mathbf{x}_5$ and compare the resulting approximations to the exact values and to the approximations obtained in Example 2.

Solution We leave it for you to confirm that

$$A\mathbf{x}_0 = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad \mathbf{x}_1 = \frac{A\mathbf{x}_0}{\max(A\mathbf{x}_0)} = \frac{1}{3}\begin{bmatrix} 3 \\ 2 \end{bmatrix} \approx \begin{bmatrix} 1.00000 \\ 0.66667 \end{bmatrix}$$

$$A\mathbf{x}_1 \approx \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1.00000 \\ 0.66667 \end{bmatrix} \approx \begin{bmatrix} 4.33333 \\ 4.00000 \end{bmatrix} \quad \mathbf{x}_2 = \frac{A\mathbf{x}_1}{\max(A\mathbf{x}_1)} \approx \frac{1}{4.33333}\begin{bmatrix} 4.33333 \\ 4.00000 \end{bmatrix} \approx \begin{bmatrix} 1.00000 \\ 0.92308 \end{bmatrix}$$

$$A\mathbf{x}_2 \approx \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1.00000 \\ 0.92308 \end{bmatrix} \approx \begin{bmatrix} 4.84615 \\ 4.76923 \end{bmatrix} \quad \mathbf{x}_3 = \frac{A\mathbf{x}_2}{\max(A\mathbf{x}_2)} \approx \frac{1}{4.84615}\begin{bmatrix} 4.84615 \\ 4.76923 \end{bmatrix} \approx \begin{bmatrix} 1.00000 \\ 0.98413 \end{bmatrix}$$

$$A\mathbf{x}_3 \approx \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1.00000 \\ 0.98413 \end{bmatrix} \approx \begin{bmatrix} 4.96825 \\ 4.95238 \end{bmatrix} \quad \mathbf{x}_4 = \frac{A\mathbf{x}_3}{\max(A\mathbf{x}_3)} \approx \frac{1}{4.96825}\begin{bmatrix} 4.96825 \\ 4.95238 \end{bmatrix} \approx \begin{bmatrix} 1.00000 \\ 0.99681 \end{bmatrix}$$

$$A\mathbf{x}_4 \approx \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1.00000 \\ 0.99681 \end{bmatrix} \approx \begin{bmatrix} 4.99361 \\ 4.99042 \end{bmatrix} \quad \mathbf{x}_5 = \frac{A\mathbf{x}_4}{\max(A\mathbf{x}_4)} \approx \frac{1}{4.99361}\begin{bmatrix} 4.99361 \\ 4.99042 \end{bmatrix} \approx \begin{bmatrix} 1.00000 \\ 0.99936 \end{bmatrix}$$

$$\lambda^{(1)} = \frac{A\mathbf{x}_1 \cdot \mathbf{x}_1}{\mathbf{x}_1 \cdot \mathbf{x}_1} = \frac{(A\mathbf{x}_1)^T\mathbf{x}_1}{\mathbf{x}_1^T\mathbf{x}_1} \approx \frac{7.00000}{1.44444} \approx 4.84615$$

$$\lambda^{(2)} = \frac{A\mathbf{x}_2 \cdot \mathbf{x}_2}{\mathbf{x}_2 \cdot \mathbf{x}_2} = \frac{(A\mathbf{x}_2)^T\mathbf{x}_2}{\mathbf{x}_2^T\mathbf{x}_2} \approx \frac{9.24852}{1.85207} \approx 4.99361$$

$$\lambda^{(3)} = \frac{A\mathbf{x}_3 \cdot \mathbf{x}_3}{\mathbf{x}_3 \cdot \mathbf{x}_3} = \frac{(A\mathbf{x}_3)^T\mathbf{x}_3}{\mathbf{x}_3^T\mathbf{x}_3} \approx \frac{9.84203}{1.96851} \approx 4.99974$$

$$\lambda^{(4)} = \frac{A\mathbf{x}_4 \cdot \mathbf{x}_4}{\mathbf{x}_4 \cdot \mathbf{x}_4} = \frac{(A\mathbf{x}_4)^T\mathbf{x}_4}{\mathbf{x}_4^T\mathbf{x}_4} \approx \frac{9.96808}{1.99362} \approx 4.99999$$

$$\lambda^{(5)} = \frac{A\mathbf{x}_5 \cdot \mathbf{x}_5}{\mathbf{x}_5 \cdot \mathbf{x}_5} = \frac{(A\mathbf{x}_5)^T\mathbf{x}_5}{\mathbf{x}_5^T\mathbf{x}_5} \approx \frac{9.99360}{1.99872} \approx 5.00000$$

Thus, $\lambda^{(5)}$ approximates the dominant eigenvalue correctly to five decimal places and $\mathbf{x}_5$ closely approximates the dominant eigenvector

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

that results by taking $t = 1$ in (6).

Whereas the power method with Euclidean scaling produces a sequence that approaches a *unit* dominant eigenvector, maximum entry scaling produces a sequence that approaches an eigenvector whose largest *component* is 1.

Rate of Convergence

If A is a symmetric matrix whose distinct eigenvalues can be arranged so that

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_k|$$

then the "rate" at which the Rayleigh quotients converge to the dominant eigenvalue λ_1 depends on the ratio $|\lambda_1|/|\lambda_2|$; that is, the convergence is slow when this ratio is near 1 and rapid when it is large—the greater the ratio, the more rapid the convergence. For example, if A is a 2×2 symmetric matrix, then the greater the ratio $|\lambda_1|/|\lambda_2|$, the greater the disparity between the scaling effects of λ_1 and λ_2 in Figure 9.2.1, and hence the greater the effect that multiplication by A has on pulling the iterates toward the eigenspace of λ_1. Indeed, the rapid convergence in Example 3 is due to the fact that $|\lambda_1|/|\lambda_2| = 5/1 = 5$, which is considered to be a large ratio. In cases where the ratio is close to 1, the convergence of the power method may be so slow that other methods must be used.

Stopping Procedures

If λ is the exact value of the dominant eigenvalue, and if a power method produces the approximation $\lambda^{(k)}$ at the kth iteration, then we call

$$\left|\frac{\lambda - \lambda^{(k)}}{\lambda}\right| \tag{12}$$

the ***relative error*** in $\lambda^{(k)}$. Expressed as a percentage it is called the ***percentage error*** in $\lambda^{(k)}$. For example, if $\lambda = 5$ and the approximation after three iterations is $\lambda^{(3)} = 5.1$, then

$$\text{relative error in } \lambda^{(3)} = \left|\frac{\lambda - \lambda^{(3)}}{\lambda}\right| = \left|\frac{5 - 5.1}{5}\right| = |-0.02| = 0.02$$

$$\text{percentage error in } \lambda^{(3)} = 0.02 \times 100\% = 2\%$$

In applications one usually knows the relative error E that can be tolerated in the dominant eigenvalue, so the goal is to stop computing iterates once the relative error in the approximation to that eigenvalue is less than E. However, there is a problem in computing the relative error from (12) in that the eigenvalue λ is unknown. To circumvent this problem, it is usual to estimate λ by $\lambda^{(k)}$ and stop the computations when

$$\left|\frac{\lambda^{(k)} - \lambda^{(k-1)}}{\lambda^{(k)}}\right| < E \tag{13}$$

The quantity on the left side of (13) is called the ***estimated relative error*** in $\lambda^{(k)}$ and its percentage form is called the ***estimated percentage error*** in $\lambda^{(k)}$.

EXAMPLE 4 | Estimated Relative Error

For the computations in Example 3, find the smallest value of k for which the estimated percentage error in $\lambda^{(k)}$ is less than 0.1%.

Solution The estimated percentage errors in the approximations in Example 3 are as follows:

	Approximation	Relative Error	Percentage Error
$\lambda^{(2)}$:	$\left\|\frac{\lambda^{(2)} - \lambda^{(1)}}{\lambda^{(2)}}\right\| \approx \left\|\frac{4.99361 - 4.84615}{4.99361}\right\|$	≈ 0.02953	$= 2.953\%$
$\lambda^{(3)}$:	$\left\|\frac{\lambda^{(3)} - \lambda^{(2)}}{\lambda^{(3)}}\right\| \approx \left\|\frac{4.99974 - 4.99361}{4.99974}\right\|$	≈ 0.00123	$= 0.123\%$
$\lambda^{(4)}$:	$\left\|\frac{\lambda^{(4)} - \lambda^{(3)}}{\lambda^{(4)}}\right\| \approx \left\|\frac{4.99999 - 4.99974}{4.99999}\right\|$	≈ 0.00005	$= 0.005\%$
$\lambda^{(5)}$:	$\left\|\frac{\lambda^{(5)} - \lambda^{(4)}}{\lambda^{(5)}}\right\| \approx \left\|\frac{5.00000 - 4.99999}{5.00000}\right\|$	≈ 0.00000	$= 0\%$

Thus, $\lambda^{(4)} = 4.99999$ is the first approximation whose estimated percentage error is less than 0.1%.

Remark A rule for terminating an iterative process is called a ***stopping procedure***. In the exercises, we will discuss stopping procedures for the power method that are based on the dominant eigenvector rather than the dominant eigenvalue.

Exercise Set 9.2

In Exercises 1–2, the distinct eigenvalues of a matrix are given. Determine whether A has a dominant eigenvalue, and if so, find it.

1. **a.** $\lambda_1 = 7,\ \lambda_2 = 3,\ \lambda_3 = -8,\ \lambda_4 = 1$

b. $\lambda_1 = -5,\ \lambda_2 = 3,\ \lambda_3 = 2,\ \lambda_4 = 5$

2. **a.** $\lambda_1 = 1,\ \lambda_2 = 0,\ \lambda_3 = -3,\ \lambda_4 = 2$

b. $\lambda_1 = -3,\ \lambda_2 = -2,\ \lambda_3 = -1,\ \lambda_4 = 3$

In Exercises 3–4, apply the power method with Euclidean scaling to the matrix A, starting with $\mathbf{x}_0$ and stopping at $\mathbf{x}_4$. Compare the resulting approximations to the exact values of the dominant eigenvalue and the corresponding unit eigenvector.

3. $A = \begin{bmatrix} 5 & -1 \\ -1 & -1 \end{bmatrix};\ \mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

4. $A = \begin{bmatrix} 7 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 5 \end{bmatrix};\ \mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

In Exercises 5–6, apply the power method with maximum entry scaling to the matrix A, starting with $\mathbf{x}_0$ and stopping at $\mathbf{x}_4$. Compare the resulting approximations to the exact values of the dominant eigenvalue and the corresponding scaled eigenvector.

5. $A = \begin{bmatrix} 1 & -3 \\ -3 & 5 \end{bmatrix};\ \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

6. $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 4 \end{bmatrix};\ \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

7. Let

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix};\ \mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

a. Use the power method with maximum entry scaling to approximate a dominant eigenvector of A. Start with $\mathbf{x}_0$, round off all computations to three decimal places, and stop after three iterations.

b. Use the result in part (a) and the Rayleigh quotient to approximate the dominant eigenvalue of A.

c. Find the exact values of the eigenvector and eigenvalue approximated in parts (a) and (b).

d. Find the percentage error in the approximation of the dominant eigenvalue.

8. Repeat the directions of Exercise 7 with

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 10 \end{bmatrix};\ \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

In Exercises 9–10, a matrix A with a dominant eigenvalue and a sequence $\mathbf{x}_0, A\mathbf{x}_0, \ldots, A^5\mathbf{x}_0$ are given. Use Formulas (9) and (10) to approximate the dominant eigenvalue and a corresponding eigenvector.

9. $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix};\ \mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ A\mathbf{x}_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix},\ A^2\mathbf{x}_0 = \begin{bmatrix} 5 \\ 4 \end{bmatrix},$

$A^3\mathbf{x}_0 = \begin{bmatrix} 13 \\ 14 \end{bmatrix},\ A^4\mathbf{x}_0 = \begin{bmatrix} 41 \\ 40 \end{bmatrix},\ A^5\mathbf{x}_0 = \begin{bmatrix} 121 \\ 122 \end{bmatrix}$

10. $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix};\ \mathbf{x}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},\ A\mathbf{x}_0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix},\ A^2\mathbf{x}_0 = \begin{bmatrix} 4 \\ 5 \end{bmatrix},$

$A^3\mathbf{x}_0 = \begin{bmatrix} 14 \\ 13 \end{bmatrix},\ A^4\mathbf{x}_0 = \begin{bmatrix} 40 \\ 41 \end{bmatrix},\ A^5\mathbf{x}_0 = \begin{bmatrix} 122 \\ 121 \end{bmatrix}$

11. Consider matrices

$$A = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_0 = \begin{bmatrix} a \\ b \end{bmatrix}$$

where $\mathbf{x}_0$ is a unit vector and $a \neq 0$. Show that even though the matrix A is symmetric and has a dominant eigenvalue, the power sequence (1) in Theorem 9.2.1 does not converge. This shows that the requirement in that theorem that the dominant eigenvalue be positive is essential.

12. Use the power method with Euclidean scaling to approximate the dominant eigenvalue and a corresponding eigenvector of A. Choose your own starting vector, and stop when the estimated percentage error in the eigenvalue approximation is less than 0.1%.

a. $\begin{bmatrix} 1 & 3 & 3 \\ 3 & 4 & -1 \\ 3 & -1 & 10 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 2 & -1 & 1 \\ 1 & -1 & 4 & 1 \\ 1 & 1 & 1 & 8 \end{bmatrix}$

13. Repeat Exercise 12, but this time stop when all corresponding entries in two successive eigenvector approximations differ by less than 0.01 in absolute value.

14. Repeat Exercise 12 using maximum entry scaling.

Working with Proofs

15. Prove: If A is a nonzero $n \times n$ matrix, then A^TA and AA^T have positive dominant eigenvalues.

16. **(*For readers familiar with proof by induction*)** Let A be an $n \times n$ matrix, let $\mathbf{x}_0$ be a unit vector in R^n, and define the sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots$ by

$$\mathbf{x}_1 = \frac{A\mathbf{x}_0}{\|A\mathbf{x}_0\|},\quad \mathbf{x}_2 = \frac{A\mathbf{x}_1}{\|A\mathbf{x}_1\|}, \ldots,\quad \mathbf{x}_k = \frac{A\mathbf{x}_{k-1}}{\|A\mathbf{x}_{k-1}\|}, \ldots$$

Prove by induction that $\mathbf{x}_k = A^k\mathbf{x}_0 / \|A^k\mathbf{x}_0\|$.

17. **(*For readers familiar with proof by induction*)** Let A be an $n \times n$ matrix, let $\mathbf{x}_0$ be a nonzero vector in R^n, and define the sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots$ by

$$\mathbf{x}_1 = \frac{A\mathbf{x}_0}{\max(A\mathbf{x}_0)}, \quad \mathbf{x}_2 = \frac{A\mathbf{x}_1}{\max(A\mathbf{x}_1)}, \ldots, \quad \mathbf{x}_k = \frac{A\mathbf{x}_{k-1}}{\max(A\mathbf{x}_{k-1})}, \ldots$$

Prove by induction that

$$\mathbf{x}_k = \frac{A^k\mathbf{x}_0}{\max(A^k\mathbf{x}_0)}$$

Working with Technology

T1. Use your technology utility to duplicate the computations in Example 2.

T2. Use your technology utility to duplicate the computations in Example 3.

9.3 Comparison of Procedures for Solving Linear Systems

There is an old saying that "time is money." This is especially true in industry where the cost of solving a linear system is generally determined by the time it takes for a computer to perform the required computations. This typically depends both on the speed of the computer processor and on the number of operations required by the algorithm. Thus, choosing the right algorithm has important financial implication in an industrial or research setting. In this section we will discuss some of the factors that affect the choice of algorithms for solving large-scale linear systems.

Flops and the Cost of Solving a Linear System

In computer jargon, an arithmetic operation $(+, -, *, \div)$ on two real numbers is called a ***flop***, which is an acronym for "floating-point operation."* The total number of flops required to solve a problem, which is called the ***cost*** of the solution, provides a convenient way of choosing between various algorithms for solving the problem. When needed, the cost in flops can be converted to units of time or money if the speed of the computer processor and the financial aspects of its operation are known. For example, today's fastest computers are capable of performing in excess of 17 petaflops/s (1 petaflop $= 10^{15}$ flops). Thus, an algorithm that costs 1,000,000 flops would be performed in 0.000000001 second. By contrast, today's personal computers can perform in excess of 80 gigaflops/s (1 gigaflop $= 10^9$ flops). Thus, an algorithm that costs 1,000,000 flops would be performed on a personal computer in 0.0000125 second.

It is now common in computer jargon to write "FLOPs" to mean the number of "flops per second." However, we will write "flops" simply as the plural of "flop." When needed, we will write flops per second as flops/s.

To illustrate how costs (in flops) can be computed, let us count the number of flops required to solve a linear system of n equations in n unknowns by Gauss–Jordan elimination. For this purpose we will need the following formulas for the sum of the first n positive integers and the sum of the squares of the first n positive integers:

$$\boxed{1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}} \tag{1}$$

$$\boxed{1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}} \tag{2}$$

*Real numbers are stored in computers as numerical approximations called ***floating-point numbers***. In base 10, a floating-point number has the form $\pm .d_1 d_2 \cdots d_n \times 10^m$, where m is an integer, called the ***mantissa***, and n is the number of digits to the right of the decimal point. The value of n varies with the computer. In some literature the term *flop* is used as a measure of processing speed and stands for "floating-point operations *per second*." In our usage it is interpreted as a counting unit.

Let $A\mathbf{x} = \mathbf{b}$ be a linear system of n equations in n unknowns to be solved by Gauss–Jordan elimination (or, equivalently, by Gaussian elimination with back substitution). For simplicity, let us assume that A is invertible and that no row interchanges are required to reduce the augmented matrix $[A \mid \mathbf{b}]$ to row echelon form. The diagrams that accompany the following analysis provide a convenient way of counting the operations required to introduce a leading 1 in the first row and then zeros below it. In our operation counts, we will lump divisions and multiplications together as "multiplications," and we will lump additions and subtractions together as "additions."

Step 1. It requires n flops (multiplications) to introduce the leading 1 in the first row.

$$\left[\begin{array}{cccccc|c} 1 & \times & \times & \cdots & \times & \times & \times \\ \bullet & \bullet & \bullet & \cdots & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \cdots & \bullet & \bullet & \bullet \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ \bullet & \bullet & \bullet & \cdots & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \cdots & \bullet & \bullet & \bullet \end{array}\right] \qquad \left[\begin{array}{l} \times \text{ denotes a quantity that is being computed.} \\ \bullet \text{ denotes a quantity that is not being computed.} \\ \text{The augmented matrix size is } n \times (n+1). \end{array}\right]$$

Step 2. It requires n multiplications and n additions to introduce a zero below the leading 1, and there are $n - 1$ rows below the leading 1, so the number of flops required to introduce zeros below the leading 1 is $2n(n - 1)$.

$$\left[\begin{array}{cccccc|c} 1 & \bullet & \bullet & \cdots & \bullet & \bullet & \bullet \\ 0 & \times & \times & \cdots & \times & \times & \times \\ 0 & \times & \times & \cdots & \times & \times & \times \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & \times & \times & \cdots & \times & \times & \times \\ 0 & \times & \times & \cdots & \times & \times & \times \end{array}\right]$$

Column 1. Combining Steps 1 and 2, the number of flops required for column 1 is

$$n + 2n(n - 1) = 2n^2 - n$$

Column 2. The procedure for column 2 is the same as for column 1, except that now we are dealing with one less row and one less column. Thus, the number of flops required to introduce the leading 1 in row 2 and the zeros below it can be obtained by replacing n by $n - 1$ in the flop count for the first column. Thus, the number of flops required for column 2 is

$$2(n - 1)^2 - (n - 1)$$

Column 3. By the argument for column 2, the number of flops required for column 3 is

$$2(n - 2)^2 - (n - 2)$$

Total. The pattern should now be clear. The total number of flops required to create the n leading 1's and the associated zeros is

$$(2n^2 - n) + [2(n - 1)^2 - (n - 1)] + [2(n - 2)^2 - (n - 2)] + \cdots + (2 - 1)$$

which we can rewrite as

$$2[n^2 + (n - 1)^2 + \cdots + 1] - [n + (n - 1) + \cdots + 1]$$

or on applying Formulas (1) and (2) as

$$2\left[\frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)}{2}\right] = \frac{2}{3}n^3 + \frac{1}{2}n^2 - \frac{1}{6}n$$

Next, let us count the number of operations required to complete the backward phase (the back substitution).

Column n. It requires $n-1$ multiplications and $n-1$ additions to introduce zeros above the leading 1 in the nth column, so the total number of flops required for the column is $2(n-1)$.

$$\left[\begin{array}{cccccc|c} 1 & \bullet & \bullet & \cdots & \bullet & 0 & \times \\ 0 & 1 & \bullet & \cdots & \bullet & 0 & \times \\ 0 & 0 & 1 & \cdots & \bullet & 0 & \times \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & \times \\ 0 & 0 & 0 & \cdots & 0 & 1 & \bullet \end{array}\right]$$

Column $(n-1)$. The procedure is the same as for Step 1, except that now we are dealing with one less row. Thus, the number of flops required for the $(n-1)$st column is $2(n-2)$.

$$\left[\begin{array}{cccccc|c} 1 & \bullet & \bullet & \cdots & 0 & 0 & \times \\ 0 & 1 & \bullet & \cdots & 0 & 0 & \times \\ 0 & 0 & 1 & \cdots & 0 & 0 & \times \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & \bullet \\ 0 & 0 & 0 & \cdots & 0 & 1 & \bullet \end{array}\right]$$

Column $(n-2)$. By the argument for column $(n-1)$, the number of flops required for column $(n-2)$ is $2(n-3)$.

Total. The pattern should now be clear. The total number of flops to complete the backward phase is

$$2(n-1)+2(n-2)+2(n-3)+\cdots+2(n-n)=2[n^2-(1+2+\cdots+n)]$$

which we can rewrite using Formula (1) as

$$2\left(n^2-\frac{n(n+1)}{2}\right)=n^2-n$$

In summary, we have shown that for Gauss–Jordan elimination the number of flops required for the forward and backward phases is

$$\text{flops for forward phase} = \tfrac{2}{3}n^3+\tfrac{1}{2}n^2-\tfrac{1}{6}n \tag{3}$$

$$\text{flops for backward phase} = n^2-n \tag{4}$$

Thus, the total cost of solving a linear system by Gauss–Jordan elimination is

$$\text{flops for both phases} = \tfrac{2}{3}n^3+\tfrac{3}{2}n^2-\tfrac{7}{6}n \tag{5}$$

Cost Estimates for Solving Large Linear Systems

It is a property of polynomials that for large values of the independent variable the term of highest power makes the major contribution to the value of the polynomial. Thus, for *large* linear systems we can use (3) and (4) to approximate the number of flops in the forward and backward phases as

$$\text{flops for forward phase} \approx \tfrac{2}{3}n^3 \tag{6}$$

$$\text{flops for backward phase} \approx n^2 \tag{7}$$

This shows that it is more costly to execute the forward phase than the backward phase for large linear systems. Indeed, the cost difference between the forward and backward phases can be enormous, as the next example shows.

We leave it as an exercise for you to confirm the results in **Table 1**.

TABLE 1

Approximate Cost for an $n \times n$ Matrix A with Large n	
Algorithm	**Cost in Flops**
Gauss–Jordan elimination (forward phase)	$\approx \frac{2}{3}n^3$
Gauss–Jordan elimination (backward phase)	$\approx n^2$
LU-decomposition of A	$\approx \frac{2}{3}n^3$
Forward substitution to solve $L\mathbf{y} = \mathbf{b}$	$\approx n^2$
Backward substitution to solve $U\mathbf{x} = \mathbf{y}$	$\approx n^2$
A^{-1} by reducing $[A \mid I]$ to $[I \mid A^{-1}]$	$\approx 2n^3$
Compute $A^{-1}\mathbf{b}$	$\approx 2n^3$

The cost in flops for Gaussian elimination is the same as that for the forward phase of Gauss–Jordan elimination.

EXAMPLE 1 | Cost of Solving a Large Linear System

Approximate the time required to execute the forward and backward phases of Gauss–Jordan elimination for a system of one million ($= 10^6$) equations in one million unknowns using a computer that can execute 10 petaflops per second (1 petaflop $= 10^{15}$ flops).

Solution We have $n = 10^6$ for the given system, so from (6) and (7) the number of petaflops required for the forward and backward phases is

$$\text{petaflops for forward phase} \approx \tfrac{2}{3}n^3 \times 10^{-15} = \tfrac{2}{3}(10^6)^3 \times 10^{-15} = \tfrac{2}{3} \times 10^3$$
$$\text{petaflops for backward phase} \approx n^2 \times 10^{-15} = (10^6)^2 \times 10^{-15} = 10^{-3}$$

Thus, at 10 petaflops/s the execution times for the forward and backward phases are

$$\text{time for forward phase} \approx \left(\tfrac{2}{3} \times 10^3\right) \times 10^{-1}\ \text{s} \approx 66.67\ \text{s}$$
$$\text{time for backward phase} \approx (10^{-3}) \times 10^{-1}\ \text{s} \approx 0.0001\ \text{s}$$

Considerations in Choosing an Algorithm for Solving a Linear System

For a *single* linear system $A\mathbf{x} = \mathbf{b}$ of n equations in n unknowns, the methods of LU-decomposition and Gauss–Jordan elimination differ in bookkeeping but otherwise involve the same number of flops. Thus, neither method has a cost advantage over the other. However, LU-decomposition has the following advantages that make it the method of choice:

- Gauss–Jordan elimination and Gaussian elimination both use the augmented matrix $[A \mid \mathbf{b}]$, so $\mathbf{b}$ must be known. In contrast, LU-decomposition uses only the matrix A, so once that decomposition is known it can be used with as many right-hand sides as are required.
- The LU-decomposition that is computed to solve $A\mathbf{x} = \mathbf{b}$ can be used to compute A^{-1}, if needed, with little additional work.
- For large linear systems in which computer memory is at a premium, one can dispense with the storage of the 1's and zeros that appear on or below the main diagonal of U, since those entries are known from the form of U. The space that this opens up

can then be used to store the entries of L, thereby reducing the amount of memory required to solve the system.

- If A is a large matrix consisting mostly of zeros, and if the nonzero entries are concentrated in a "band" around the main diagonal, then there are techniques that can be used to reduce the cost of LU-decomposition, giving it an advantage over Gauss–Jordan elimination.

Exercise Set 9.3

1. A certain computer can execute 10 gigaflops per second. Use Formula (5) to find the time required to solve the system using Gauss–Jordan elimination.

a. A system of 1000 equations in 1000 unknowns.

b. A system of 10,000 equations in 10,000 unknowns.

c. A system of 100,000 equations in 100,000 unknowns.

2. A certain computer can execute 100 gigaflops per second. Use Formula (5) to find the time required to solve the system using Gauss–Jordan elimination.

a. A system of 10,000 equations in 10,000 unknowns.

b. A system of 100,000 equations in 100,000 unknowns.

c. A system of 1,000,000 equations in 1,000,000 unknowns.

3. A certain computer can execute 70 gigaflops per second. Use **Table 1** to estimate the time required to perform the following operations on the invertible $10{,}000 \times 10{,}000$ matrix A.

a. Execute the forward phase of Gauss–Jordan elimination.

b. Execute the backward phase of Gauss–Jordan elimination.

c. LU-decomposition of A.

d. Find A^{-1} by reducing $[A \mid I]$ to $[I \mid A^{-1}]$.

4. The IBM Sequoia computer can operate at speeds in excess of 16 petaflops per second (1 petaflop $= 10^{15}$ flops). Use Table 1 to estimate the time required to perform the following operations on an invertible $100{,}000 \times 100{,}000$ matrix A.

a. Execute the forward phase of Gauss–Jordan elimination.

b. Execute the backward phase of Gauss–Jordan elimination.

c. LU-decomposition of A.

d. Find A^{-1} by reducing $[A \mid I]$ to $[I \mid A^{-1}]$.

5. **a.** Approximate the time required to execute the forward phase of Gauss–Jordan elimination for a system of 100,000 equations in 100,000 unknowns using a computer that can execute 1 gigaflop per second. Do the same for the backward phase. (See Table 1.)

b. How many gigaflops per second must a computer be able to execute to find the LU-decomposition of a matrix of size $10{,}000 \times 10{,}000$ in less than 0.5 s? (See Table 1.)

6. About how many teraflops per second must a computer be able to execute to find the inverse $100{,}000 \times 100{,}000$ matrix in less than 0.5 s? (1 teraflop $= 10^{12}$ flops.)

In Exercises 7–10, suppose A and B are $n \times n$ matrices and c is a real number.

7. How many flops are required to compute cA?

8. How many flops are required to compute $A + B$?

9. How many flops are required to compute AB?

10. If A is a diagonal matrix and k is a positive integer, how many flops are required to compute A^k?

9.4 Singular Value Decomposition

In this section we will discuss an extension of the diagonalization theory for $n \times n$ symmetric matrices to general $m \times n$ matrices. The results that we will develop in this section have applications to compression, storage, and transmission of digitized information and form the basis for many of the best computational algorithms that are currently available for solving linear systems.

Decompositions of Square Matrices

We saw in Formula (2) of Section 7.2 that every symmetric matrix A with real entries can be expressed as

$$A = PDP^T \tag{1}$$

where P is an orthogonal matrix whose columns are eigenvectors of A, and D is the diagonal matrix whose diagonal entries are the eigenvalues corresponding to the column vectors of P. In this section we will call (1) an ***eigenvalue decomposition*** of A (abbreviated EVD of A).

If an $n \times n$ matrix A is not symmetric, then it does not have an eigenvalue decomposition, but it does have a ***Hessenberg decomposition***

$$A = PHP^T$$

in which P is an orthogonal matrix and H is in upper Hessenberg form (Theorem 7.2.4). Moreover, if A has real eigenvalues, then it has a ***Schur decomposition***

$$A = PSP^T$$

in which P is an orthogonal matrix and S is upper triangular (Theorem 7.2.3).

The eigenvalue, Hessenberg, and Schur decompositions are important in numerical algorithms not only because the matrices D, H, and S have simpler forms than A, but also because the orthogonal matrices that appear in these factorizations do not magnify roundoff error. To see why this is so, suppose that $\hat{\mathbf{x}}$ is a column vector whose entries are known exactly and that

$$\mathbf{x} = \hat{\mathbf{x}} + \mathbf{e}$$

is the vector that results when roundoff error is present in the entries of $\hat{\mathbf{x}}$. If P is an orthogonal matrix, then the length-preserving property of orthogonal transformations implies that

$$\|P\mathbf{x} - P\hat{\mathbf{x}}\| = \|\mathbf{x} - \hat{\mathbf{x}}\| = \|\mathbf{e}\|$$

which tells us that the error in approximating $P\hat{\mathbf{x}}$ by $P\mathbf{x}$ has the same magnitude as the error in approximating $\hat{\mathbf{x}}$ by $\mathbf{x}$.

There are two main paths that one might follow in looking for other kinds of decompositions of a general square matrix A: One might look for decompositions of the form

$$A = PJP^{-1}$$

in which P is invertible but not necessarily orthogonal, or one might look for decompositions of the form

$$A = U\Sigma V^T$$

in which U and V are orthogonal but not necessarily the same. The first path leads to decompositions in which J is either diagonal or a certain kind of block diagonal matrix, called a ***Jordan canonical form*** in honor of the French mathematician Camille Jordan (see p. 538). Jordan canonical forms, which we will not consider in this text, are important theoretically and in certain applications, but they are of lesser importance numerically because of the roundoff difficulties that result from the lack of orthogonality in P. In this section we will focus on the second path.

Singular Values

Since matrix products of the form A^TA will play an important role in our work, we will begin with two basic theorems about them.

Theorem 9.4.1

If A is an $m \times n$ matrix, then:

(*a*) A and A^TA have the same null space.

(*b*) A and A^TA have the same row space.

(*c*) A^T and A^TA have the same column space.

(*d*) A and A^TA have the same rank.

We will prove part (*a*) and leave the remaining proofs for the exercises.

Proof (a) We must show that every solution of $A\mathbf{x} = \mathbf{0}$ is a solution of $A^TA\mathbf{x} = \mathbf{0}$, and conversely. If $\mathbf{x}_0$ is any solution of $A\mathbf{x} = \mathbf{0}$, then $\mathbf{x}_0$ is also a solution of $A^TA\mathbf{x} = \mathbf{0}$ since

$$A^TA\mathbf{x}_0 = A^T(A\mathbf{x}_0) = A^T\mathbf{0} = \mathbf{0}$$

Conversely, if $\mathbf{x}_0$ is any solution of $A^TA\mathbf{x} = \mathbf{0}$, then $\mathbf{x}_0$ is in the null space of A^TA and hence is orthogonal to all vectors in the row space of A^TA by part (s) of Theorem 8.2.4. However, A^TA is symmetric, so $\mathbf{x}_0$ is also orthogonal to every vector in the column space of A^TA. In particular, $\mathbf{x}_0$ must be orthogonal to the vector $(A^TA)\mathbf{x}_0$; that is,

$$\mathbf{x}_0 \cdot (A^TA)\mathbf{x}_0 = 0$$

Using the first row in Table 1 of Section 3.2 and properties of the transpose operation we can rewrite this as

$$\mathbf{x}_0^T(A^TA)\mathbf{x}_0 = (A\mathbf{x}_0)^T(A\mathbf{x}_0) = (A\mathbf{x}_0)\cdot(A\mathbf{x}_0) = \|A\mathbf{x}_0\|^2 = 0$$

which implies that $A\mathbf{x}_0 = \mathbf{0}$, thereby proving that $\mathbf{x}_0$ is a solution of $A\mathbf{x}_0 = \mathbf{0}$. ■

Theorem 9.4.2

If A is an $m \times n$ matrix, then:

(a) A^TA is orthogonally diagonalizable.

(b) The eigenvalues of A^TA are nonnegative real numbers.

Proof (a) The matrix A^TA, being symmetric, is orthogonally diagonalizable by Theorem 7.2.1.

Proof (b) Since A^TA is orthogonally diagonalizable, there is an orthonormal basis for R^n consisting of eigenvectors of A^TA, say $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. If we let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the corresponding eigenvalues, then for $1 \le i \le n$ we have

$$\begin{aligned} \|A\mathbf{v}_i\|^2 &= A\mathbf{v}_i \cdot A\mathbf{v}_i = \mathbf{v}_i \cdot A^TA\mathbf{v}_i \qquad \text{[Formula (26) of Section 3.2]} \\ &= \mathbf{v}_i \cdot \lambda_i\mathbf{v}_i = \lambda_i(\mathbf{v}_i\cdot\mathbf{v}_i) = \lambda_i\|\mathbf{v}_i\|^2 = \lambda_i \end{aligned}$$

It follows from this relationship that $\lambda_i \ge 0$. ■

We will assume throughout this section that the eigenvalues of A^TA are named so that

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n \ge 0$$

and hence that

$$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n \ge 0$$

Definition 1

If A is an $m \times n$ matrix, and if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A^TA, then the numbers

$$\sigma_1 = \sqrt{\lambda_1}, \quad \sigma_2 = \sqrt{\lambda_2}, \ldots, \quad \sigma_n = \sqrt{\lambda_n}$$

are called the ***singular values*** of A.

EXAMPLE 1 | Singular Values

Find the singular values of the matrix

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

Solution The first step is to find the eigenvalues of the matrix

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

The characteristic polynomial of A^TA is

$$\lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1)$$

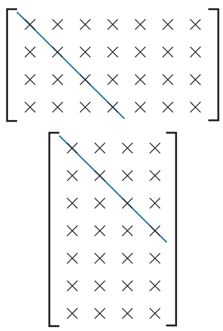
so the eigenvalues of A^TA are $\lambda_1 = 3$ and $\lambda_2 = 1$ and the singular values of A in order of decreasing size are

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{3}, \quad \sigma_2 = \sqrt{\lambda_2} = 1$$

Singular Value Decomposition

Before turning to the main result in this section, we will find it useful to extend the notion of a "main diagonal" to matrices that are not square. We define the ***main diagonal*** of an $m \times n$ matrix to be the line of entries shown in **Figure 9.4.1**—it starts at the upper left corner and extends diagonally as far as it can go. We will refer to the entries on the main diagonal as the ***diagonal entries***.

We are now ready to consider the main result in this section, which is concerned with a specific way of factoring a general $m \times n$ matrix A. This factorization, called ***singular value decomposition*** (abbreviated SVD) will be given in two forms, a brief form that captures the main idea, and an expanded form that spells out the details. The proof is given at the end of this section.



Main diagonals

FIGURE 9.4.1

Theorem 9.4.3

Singular Value Decomposition (Brief Form)

If A is an $m \times n$ matrix of rank k, then A can be expressed in the form $A = U\Sigma V^T$, where Σ has size $m \times n$ and can be expressed in partitioned form as

$$\Sigma = \left[\begin{array}{c|c} D & 0_{k\times(n-k)} \\ \hline 0_{(m-k)\times k} & 0_{(m-k)\times(n-k)} \end{array}\right]$$

in which D is a diagonal $k \times k$ matrix whose successive entries are the first k singular values of A in nonincreasing order, U is an $m \times n$ orthogonal matrix, and V is an $n \times n$ orthogonal matrix.

Historical Note

Harry Bateman (1882–1946)

The term *singular value* is apparently due to the British-born mathematician Harry Bateman, who used it in a research paper published in 1908. Bateman emigrated to the United States in 1910, teaching at Bryn Mawr College, Johns Hopkins University, and finally at the California Institute of Technology. Interestingly, he was awarded his Ph.D. in 1913 by Johns Hopkins at which point in time he was already an eminent mathematician with 60 publications to his name.

[Image: Courtesy of the Archives, California Institute of Technology]

Theorem 9.4.4

Singular Value Decomposition (Expanded Form)

If A is an $m \times n$ matrix of rank k, then A can be factored as

$$A = U\Sigma V^T = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_k \mid \mathbf{u}_{k+1} \quad \cdots \quad \mathbf{u}_m] \left[\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & \\ 0 & \sigma_2 & \cdots & 0 & 0_{k\times(n-k)} \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & \sigma_k & \\ \hline & & 0_{(m-k)\times k} & & 0_{(m-k)\times(n-k)} \end{array}\right] \left[\begin{array}{c} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \\ \hline \mathbf{v}_{k+1}^T \\ \vdots \\ \mathbf{v}_n^T \end{array}\right]$$

in which U, Σ, and V have sizes $m \times m$, $m \times n$, and $n \times n$, respectively, and in which:

(a) $V = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n]$ orthogonally diagonalizes A^TA.

(b) The nonzero diagonal entries of Σ are $\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2}, \ldots, \sigma_k = \sqrt{\lambda_k}$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the nonzero eigenvalues of A^TA corresponding to the column vectors of V.

(c) The column vectors of V are ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$.

(d) $\mathbf{u}_i = \dfrac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \dfrac{1}{\sigma_i}A\mathbf{v}_i \qquad (i = 1, 2, \ldots, k)$

(e) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is an orthonormal basis for col(A).

(f) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ is an extension of $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ to an orthonormal basis for R^m.

The vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are called the ***left singular vectors*** of A, and the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are called the ***right singular vectors*** of A.

EXAMPLE 2 | Singular Value Decomposition if A Is Not Square

Find a singular value decomposition of the matrix

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

Solution We showed in Example 1 that the eigenvalues of A^TA are $\lambda_1 = 3$ and $\lambda_2 = 1$ and that the corresponding singular values of A are $\sigma_1 = \sqrt{3}$ and $\sigma_2 = 1$. We leave it for you to verify that

$$\mathbf{v}_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$$

are eigenvectors corresponding to λ_1 and λ_2, respectively, and that $V = [\mathbf{v}_1 \mid \mathbf{v}_2]$ orthogonally diagonalizes A^TA. From part (d) of Theorem 9.4.4, the vectors

$$\mathbf{u}_1 = \frac{1}{\sigma_1}A\mathbf{v}_1 = \frac{\sqrt{3}}{3}\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{6} \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2}A\mathbf{v}_2 = (1)\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$$

are two of the three column vectors of U. Note that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthonormal, as expected. We could extend the set $\{\mathbf{u}_1, \mathbf{u}_2\}$ to an orthonormal basis for R^3. However, the computations will be easier if we first remove the messy radicals by multiplying $\mathbf{u}_1$ and $\mathbf{u}_2$ by appropriate scalars. Thus, we will look for a unit vector $\mathbf{u}_3$ that is orthogonal to

$$\sqrt{6}\,\mathbf{u}_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \sqrt{2}\,\mathbf{u}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

To satisfy these two orthogonality conditions, the vector $\mathbf{u}_3$ must be a solution of the homogeneous linear system

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

We leave it for you to show that a general solution of this system is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

Normalizing the vector on the right yields

$$\mathbf{u}_3 = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

Thus, the singular value decomposition of A is

$$\underset{A}{\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}} = \underset{U}{\begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}} \underset{V^T}{\begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}}$$

You may want to confirm the validity of this equation by multiplying out the matrices on the right side.

OPTIONAL: We conclude this section with an optional proof of Theorem 9.4.4.

Proof of Theorem 9.4.4 For notational simplicity we will prove this theorem in the case where A is an $n \times n$ matrix. To modify the argument for an $m \times n$ matrix you need only make the notational adjustments required to account for the possibility that $m > n$ or $n > m$.

The matrix A^TA is symmetric, so it has an eigenvalue decomposition

$$A^TA = VDV^T$$

in which the column vectors of

$$V = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_n]$$

are unit eigenvectors of A^TA, and D is a diagonal matrix whose successive diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A^TA corresponding in succession to the column vectors of V. Since A is assumed to have rank k, it follows from Theorem 9.4.1 that A^TA also has rank k. It follows as well that D has rank k, since it is similar to A^TA and rank is a similarity invariant. Thus, the diagonal matrix D can be expressed in the form

$$D = \begin{bmatrix} \lambda_1 & & & & & & 0 \\ & \lambda_2 & & & & & \\ & & \ddots & & & & \\ & & & \lambda_k & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 0 \end{bmatrix} \tag{2}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$. Now let us consider the set of image vectors

$$\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_n\} \tag{3}$$

This is an orthogonal set, for if $i \neq j$, then the orthogonality of $\mathbf{v}_i$ and $\mathbf{v}_j$ implies that

$$A\mathbf{v}_i \cdot A\mathbf{v}_j = \mathbf{v}_i \cdot A^TA\mathbf{v}_j = \mathbf{v}_i \cdot \lambda_j\mathbf{v}_j = \lambda_j(\mathbf{v}_i \cdot \mathbf{v}_j) = 0$$

Moreover, the first k vectors in (3) are nonzero since we showed in the proof of Theorem 9.4.2(b) that $\|A\mathbf{v}_i\|^2 = \lambda_i$ for $i = 1, 2, \ldots, n$, and we have assumed that the first k diagonal entries in (2) are positive. Thus,

$$S = \{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_k\}$$

is an orthogonal set of *nonzero* vectors in the column space of A. But the column space of A has dimension k since

$$\operatorname{rank}(A) = \operatorname{rank}(A^TA) = k$$

and hence S, being a linearly independent set of k vectors, must be an orthogonal basis for col(A). If we now normalize the vectors in S, we will obtain an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ for col(A) in which

$$\mathbf{u}_i = \frac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \frac{1}{\sqrt{\lambda_i}}A\mathbf{v}_i \qquad (1 \leq i \leq k)$$

or, equivalently, in which

$$A\mathbf{v}_1 = \sqrt{\lambda_1}\mathbf{u}_1 = \sigma_1\mathbf{u}_1, \quad A\mathbf{v}_2 = \sqrt{\lambda_2}\mathbf{u}_2 = \sigma_2\mathbf{u}_2, \ldots, \quad A\mathbf{v}_k = \sqrt{\lambda_k}\mathbf{u}_k = \sigma_k\mathbf{u}_k \tag{4}$$

It follows from Theorem 6.3.6 that we can extend this to an orthonormal basis

$$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$$

for R^n. Now let U be the orthogonal matrix

$$U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{u}_{k+1} & \cdots & \mathbf{u}_n \end{bmatrix}$$

and let Σ be the diagonal matrix

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & & 0 \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_k & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 0 \end{bmatrix}$$

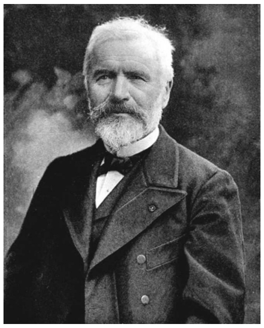
It follows from (4), and the fact that $A\mathbf{v}_i = 0$ for $i > k$, that

$$\begin{aligned} U\Sigma &= [\sigma_1\mathbf{u}_1 \quad \sigma_2\mathbf{u}_2 \quad \cdots \quad \sigma_k\mathbf{u}_k \quad 0 \quad \cdots \quad 0] \\ &= [A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad \cdots \quad A\mathbf{v}_k \quad A\mathbf{v}_{k+1} \quad \cdots \quad A\mathbf{v}_n] \\ &= AV \end{aligned}$$

which we can rewrite using the orthogonality of V as $A = U\Sigma V^T$. ■

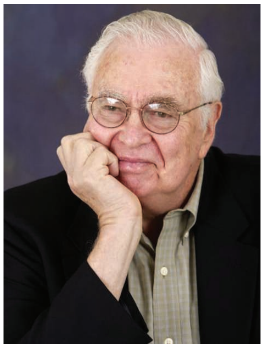
Historical Note

Eugenio Beltrami (1835–1900)



Camille Jordan (1838–1922)

Herman Klaus Weyl (1885–1955)



Gene H. Golub (1932–2007)

The theory of singular value decompositions can be traced back to the work of five people: the Italian mathematician Eugenio Beltrami, the French mathematician Camille Jordan, the English mathematician James Sylvester (see p. 36), and the German mathematicians Erhard Schmidt (see p. 369) and the mathematician Herman Weyl. More recently, the pioneering efforts of the American mathematician Gene Golub produced a stable and efficient algorithm for computing it. Beltrami and Jordan were the progenitors of the decomposition—Beltrami gave a proof of the result for real, invertible matrices with distinct singular values in 1873. Subsequently, Jordan refined the theory and eliminated the unnecessary restrictions imposed by Beltrami. Sylvester, apparently unfamiliar with the work of Beltrami and Jordan, rediscovered the result in 1889 and suggested its importance. Schmidt was the first person to show that the singular value decomposition could be used to approximate a matrix by another matrix with lower rank, and, in so doing, he transformed it from a mathematical curiosity to an important practical tool. Weyl showed how to find the lower rank approximations in the presence of error.

[Images: http://www-history.mcs.st-andrews.ac.uk/history/PictDisplay/Beltrami.html (Beltrami); The Granger Collection, New York (Jordan); Courtesy Electronic Publishing Services, Inc., New York City (Weyl); Courtesy of Hector Garcia-Molina (Golub)]

Exercise Set 9.4

*In Exercises **1–4**, find the distinct singular values of A.*

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

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

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

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

*In Exercises **5–12**, find a singular value decomposition of A.*

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

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

7. $A = \begin{bmatrix} 4 & 6 \\ 0 & 4 \end{bmatrix}$

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

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

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

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

12. $A = \begin{bmatrix} 6 & 4 \\ 0 & 0 \\ 4 & 0 \end{bmatrix}$

Working with Proofs

13. Prove: If A is an $m \times n$ matrix, then A^TA and AA^T have the same rank.

14. Prove part (d) of Theorem 9.4.1 by using part (a) of the theorem and the fact that A and A^TA have n columns.

15. **a.** Prove part (b) of Theorem 9.4.1 by first showing that $\text{row}(A^TA)$ is a subspace of $\text{row}(A)$.

b. Prove part (c) of Theorem 9.4.1 by using part (b).

16. Let $T : R^n \to R^m$ be a linear transformation whose standard matrix A has the singular value decomposition $A = U\Sigma V^T$, and let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $B' = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be the column vectors of V and U, respectively. Prove that

$$\Sigma = [T]_{B',B}$$

17. Prove that the singular values of A^TA are the squares of the singular values of A.

18. Prove that if $A = U\Sigma V^T$ is a singular value decomposition of A, then U orthogonally diagonalizes AA^T.

19. A ***polar decomposition*** of an $n \times n$ matrix A is a factorization $A = PQ$ in which P is a positive semidefinite $n \times n$ matrix with the same rank as A, and Q is an orthogonal $n \times n$ matrix.

a. Prove that if $A = U\Sigma V^T$ is the singular value decomposition of A, then $A = (U\Sigma U^T)(UV^T)$ is a polar decomposition of A.

b. Find a polar decomposition of the matrix in Exercise 5.

True-False Exercises

TF. In parts (a)–(g) determine whether the statement is true or false, and justify your answer.

a. If A is an $m \times n$ matrix, then A^TA is an $m \times m$ matrix.

b. If A is an $m \times n$ matrix, then A^TA is a symmetric matrix.

c. If A is an $m \times n$ matrix, then the eigenvalues of A^TA are positive real numbers.

d. If A is an $n \times n$ matrix, then A is orthogonally diagonalizable.

e. If A is an $m \times n$ matrix, then A^TA is orthogonally diagonalizable.

f. The eigenvalues of A^TA are the singular values of A.

g. Every $m \times n$ matrix has a singular value decomposition.

Working with Technology

T1. Use your technology utility to duplicate the computations in Example 2.

T2. For the given matrix A, use the steps in Example 2 to find matrices U, Σ, and V^T in a singular value decomposition $A = U\Sigma V^T$.

a. $A = \begin{bmatrix} -2 & -1 & 2 \\ 2 & 1 & -2 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 1 \end{bmatrix}$

9.5 Data Compression Using Singular Value Decomposition

Efficient transmission and storage of large quantities of digital data has become a major problem in our technological world. In this section we will discuss the role that singular value decomposition plays in compressing digital data so that it can be transmitted more rapidly and stored in less space. We assume here that you have read Section 9.4.

Reduced Singular Value Decomposition

Algebraically, the zero rows and columns of the matrix Σ in Theorem 9.4.4 are superfluous and can be eliminated by multiplying out the expression $U\Sigma V^T$ using block multiplication and the partitioning shown in that formula. The products that involve zero blocks as factors drop out, leaving

$$\boxed{A = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_k] \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_k \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \end{bmatrix}} \tag{1}$$

which is called a ***reduced singular value decomposition*** of A. In this text we will denote the matrices on the right side of (1) by U_1, Σ_1, and V_1^T, respectively, and we will write this equation as

$$\boxed{A = U_1\Sigma_1V_1^T} \tag{2}$$

Note that the sizes of U_1, Σ_1, and V_1^T are $m \times k$, $k \times k$, and $k \times n$, respectively, and that the matrix Σ_1 is invertible since its diagonal entries are positive.

If we multiply out on the right side of (1) using the column-row rule, then we obtain

$$\boxed{A = \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T + \cdots + \sigma_k\mathbf{u}_k\mathbf{v}_k^T} \tag{3}$$

which is called a ***reduced singular value expansion*** of A. This result applies to *all* matrices, whereas the spectral decomposition [Formula (7) of Section 7.2] applies only to symmetric matrices.

Remark It can be proved that an $m \times n$ matrix M has rank 1 if and only if it can be factored as $M = \mathbf{u}\mathbf{v}^T$, where $\mathbf{u}$ is a column vector in R^m and V is a column vector in R^n. Thus, a reduced singular value decomposition expresses a matrix A of rank k as a linear combination of k rank 1 matrices.

EXAMPLE 1 | Reduced Singular Value Decomposition

Find a reduced singular value decomposition and a reduced singular value expansion of the matrix

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

Solution In Example 2 of Section 9.4 we found the singular value decomposition

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

$$A \qquad = \qquad U \qquad \qquad \Sigma \qquad \qquad V^T$$

Since A has rank 2 (verify), it follows from (1) with $k = 2$ that the reduced singular value decomposition of A corresponding to (4) is

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

This yields the reduced singular value expansion

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T = \sqrt{3} \begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{6} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} + (1) \begin{bmatrix} 0 \\ -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$$

$$= \sqrt{3} \begin{bmatrix} \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{6} & \frac{\sqrt{3}}{6} \\ \frac{\sqrt{3}}{6} & \frac{\sqrt{3}}{6} \end{bmatrix} + (1) \begin{bmatrix} 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

Note that the matrices in the expansion have rank 1, as expected.

Data Compression and Image Processing

Singular value decompositions can be used to "compress" visual information for the purpose of reducing its required storage space and speeding up its electronic transmission. The first step in compressing a visual image is to represent it as a numerical matrix from which the visual image can be recovered when needed.

For example, a black and white photograph might be scanned as a rectangular array of pixels (points) and then stored as a matrix A by assigning each pixel a numerical value in accordance with its gray level. If 256 different gray levels are used ($0 =$ white to $255 =$ black), then the entries in the matrix would be integers between 0 and 255. The image can be recovered from the matrix A by printing or displaying the pixels with their assigned gray levels.

If the matrix A has size $m \times n$, then one might store each of its mn entries individually. An alternative procedure is to compute the reduced singular value decomposition

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T \tag{5}$$

in which $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k$, and store the σ's, the $\mathbf{u}$'s, and the $\mathbf{v}$'s. When needed, the matrix A (and hence the image it represents) can be reconstructed from (5). Since each $\mathbf{u}_j$ has m entries and each $\mathbf{v}_j$ has n entries, this method requires storage space for

$$km + kn + k = k(m + n + 1)$$

numbers. Suppose, however, that the singular values $\sigma_{r+1}, \ldots, \sigma_k$ are sufficiently small that dropping the corresponding terms in (5) produces an acceptable approximation

$$\boxed{A_r = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T} \tag{6}$$

Historical Note

Original

Reconstruction

In 1924 the U.S. Federal Bureau of Investigation (FBI) began collecting fingerprints and handprints and now has more than 100 million such prints in its files. To reduce the storage cost, the FBI began working with the Los Alamos National Laboratory, the National Bureau of Standards, and other groups in 1993 to devise rank-based compression methods for storing prints in digital form. The adjacent figure shows an original fingerprint and a reconstruction from digital data that was compressed at a ratio of 26:1.

to A and the image that it represents. We call (6) the ***rank r approximation of A***. This matrix requires storage space for only

$$rm + rn + r = r(m + n + 1)$$

numbers, compared to mn numbers required for entry-by-entry storage of A. For example, the rank 100 approximation of a 1000×1000 matrix A requires storage for only

$$100(1000 + 1000 + 1) = 200{,}100$$

numbers, compared to the 1,000,000 numbers required for entry-by-entry storage of A—a compression of almost 80%.

Figure 9.5.1 shows some approximations of a digitized mandrill image obtained using (6).

Rank 4 Rank 10 Rank 20 Rank 50 Rank 128

FIGURE 9.5.1

Exercise Set 9.5

In Exercises 1–4, find a reduced singular value decomposition of A. [Note: *Each matrix appears in Exercise Set 9.4, where you were asked to find its (unreduced) singular value decomposition.*]

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

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

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

4. $A = \begin{bmatrix} 6 & 4 \\ 0 & 0 \\ 4 & 0 \end{bmatrix}$

In Exercises 5–8, find a reduced singular value expansion of A.

5. The matrix A in Exercise 1.

6. The matrix A in Exercise 2.

7. The matrix A in Exercise 3.

8. The matrix A in Exercise 4.

9. Suppose A is a 200×500 matrix. How many numbers must be stored in the rank 100 approximation of A? Compare this with the number of entries of A.

True-False Exercises

TF. In parts **(a)–(c)** determine whether the statement is true or false, and justify your answer. Assume that $U_1\Sigma_1V_1^T$ is a reduced singular value decomposition of an $m \times n$ matrix of rank k.

a. U_1 has size $m \times k$.

b. Σ_1 has size $k \times k$.

c. V_1 has size $k \times n$.

Chapter 9 Supplementary Exercises

1. Find an LU-decomposition of $A = \begin{bmatrix} -6 & 2 \\ 6 & 0 \end{bmatrix}$.

2. Find the LDU-decomposition of the matrix A in Exercise 1.

3. Find an LU-decomposition of $A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 4 & 7 \\ 1 & 3 & 7 \end{bmatrix}$.

4. Find the LDU-decomposition of the matrix A in Exercise 3.

5. Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

a. Identify the dominant eigenvalue of A and then find the corresponding dominant unit eigenvector $\mathbf{v}$ with *positive* entries.

b. Apply the power method with Euclidean scaling to A and $\mathbf{x}_0$, stopping at $\mathbf{x}_5$. Compare your value of $\mathbf{x}_5$ to the eigenvector $\mathbf{v}$ found in part (a).

c. Apply the power method with maximum entry scaling to A and $\mathbf{x}_0$, stopping at $\mathbf{x}_5$. Compare your result with the eigenvector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

6. Consider the symmetric matrix

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

Discuss the behavior of the power sequence

$$\mathbf{x}_0, \quad \mathbf{x}_1, \ldots, \quad \mathbf{x}_k, \ldots$$

with Euclidean scaling for a general *nonzero* vector $\mathbf{x}_0$. What is it about the matrix that causes the observed behavior?

7. Suppose that a symmetric matrix A has distinct eigenvalues $\lambda_1 = 8$, $\lambda_2 = 1.4$, $\lambda_3 = 2.3$, and $\lambda_4 = -8.1$. What can you say about the convergence of the Rayleigh quotients?

8. Find a singular value decomposition of $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

9. Find a singular value decomposition of $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$.

10. Find a reduced singular value decomposition and a reduced singular value expansion of the matrix A in Exercise 9.

11. Find the reduced singular value decomposition of the matrix whose singular value decomposition is

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 24 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix}$$

12. Do orthogonally similar matrices have the same singular values? Justify your answer.

13. If P is the standard matrix for the orthogonal projection of R^n onto a subspace W, what can you say about the singular values of P?

14. Prove: If A has rank 1, then there exists a scalar k such that $A^2 = kA$.

Appendix A

Working with Proofs

Linear algebra is different from other mathematics courses that you may encounter in that it is more than a collection of problem-solving techniques. Even if you learn to solve all of the computational problems in this text, you will have fallen short in your mastery of the subject. This is because innovative uses of linear algebra typically require new techniques based on an understanding of its theorems, their interrelationships, and their proofs. While it is impossible to teach you everything you will need to do proofs, this appendix will provide some guidelines that may help.

What Is a Proof?

In essence, a proof is a "convincing argument" that justifies the truth of a mathematical statement. Although what may be convincing to one person may not be convincing to another, experience has led mathematicians to establish clear standards on what is to be considered an acceptable proof and what is not. We will try to explain here some of the logical steps required of an acceptable proof.

Formality

In high-school geometry you may have been asked to prove theorems by formally listing statements on the left and justifications on the right. That level of formality is not required in linear algebra. Rather, a proof need only be an argument, written in complete sentences, that leads step by step to a logical conclusion, and in which each step is justified by referencing some statement whose validity is either self-evident or has been previously proved.

How to Read Theorems

Most theorems are of the form

$$\textit{If } H \textit{ is true, then } C \textit{ is true.} \tag{1}$$

where H is a statement called the ***hypothesis*** and C is a statement called the ***conclusion***. In formal logic one denotes a theorem of this form as

$$H \Rightarrow C \tag{2}$$

which is read, "H implies C." A statement of this type is considered to be true if the conclusion C is true in all cases where the hypothesis H is true, and it is considered to be false if there is at least one case where H is true and C is false. As an example, consider the statement

$$\textit{If } a \textit{ and } b \textit{ are both positive numbers, then } ab \textit{ is a positive number.} \tag{3}$$

In this statement,

$$H = a \textit{ and } b \textit{ are both positive numbers} \tag{4}$$

$$C = ab \textit{ is a positive number} \tag{5}$$

Statement (3) is true because C is true in all cases where H is true. On the other hand, the statement

$$\textit{If } a \textit{ and } b \textit{ are positive integers, then } \sqrt{ab} \textit{ is a positive integer.} \tag{6}$$

is not true because there exist cases where the hypothesis is true and the conclusion is false—for example, if $a = 2$ and $b = 3$.

Sometimes it is desirable to phrase statements in a negative way. For example, statement (3) can be rephrased equivalently as

$$\textit{If ab is not a positive number, then a and b are not both positive numbers.} \tag{7}$$

If we write $\sim H$ to mean that H is false and $\sim C$ to mean that C is false, then the structure of statement (7) is

$$\sim C \Rightarrow \sim H \tag{8}$$

This is called the ***contrapositive*** form of (2). It can be shown that a statement and its contrapositive are logically equivalent; that is, if the statement is true, then so is its contrapositive and vice versa.

The ***converse*** of a theorem is the statement that results when the hypothesis and conclusion are interchanged. Thus, the converse of the statement $H \Rightarrow C$ is the statement $C \Rightarrow H$. Whereas the contrapositive of a true statement must itself be true, the converse of a true statement may or may not be true. For example, the converse of the true statement (3) is the *false* statement

If ab is a positive number, then a and b are both positive numbers.

whereas the converse of the true statement

If the numbers a and b are both positive or both negative, then ab is a positive number.

is a *true* statement.

Warning Do not confuse the terms "contrapositive" and "converse."

In those special cases where a statement $H \Rightarrow C$ and its converse $C \Rightarrow H$ are both true, we say that H and C are ***equivalent*** statements. We denote this by writing

$$H \Leftrightarrow C \tag{9}$$

which is read, "H is equivalent to C" or, more commonly, "H is true if and only if C is true." For example, if a and b are real numbers, then

$$a > b \text{ if and only if } (a - b) > 0 \tag{10}$$

To prove an "if and only if" statement of form (9), you must prove both $H \Rightarrow C$ and $C \Rightarrow H$.

Equivalent statements are often phrased in other ways. For example, statement (10) might also be expressed as

If $a > b$, then $(a - b) > 0$ and conversely.

Sometimes two true statements will give you a third true statement for free. Specifically, if it is true that $H \Rightarrow C$ and $C \Rightarrow D$, then it follows that $H \Rightarrow D$ must also be true. For example, consider the following two theorems from geometry.

Theorem 1A

If opposite sides of a quadrilateral are parallel, then the quadrilateral is a parallelogram.

Theorem 2A

Opposite sides of a parallelogram have equal lengths.

Because the conclusion of the first theorem is essentially the hypothesis of the second, the two theorems together yield the following third theorem.

Theorem 3A

If opposite sides of a quadrilateral are parallel, then they have equal lengths.

To take this idea a step further, three true statements can sometimes yield three other true statements for free. Specifically, if

$$H \Rightarrow C, \quad C \Rightarrow D, \quad D \Rightarrow H \tag{11}$$

then we have the implication loop in **Figure A.1**, from which we see that

$$C \Rightarrow H, \quad D \Rightarrow C, \quad H \Rightarrow D$$

By combining this result with (11) we obtain

$$H \Leftrightarrow C, \quad C \Leftrightarrow D, \quad D \Leftrightarrow H \tag{12}$$

In summary, if you want to prove the three equivalences in (12) you need only prove the three implications in (11).

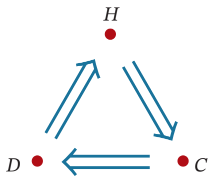


FIGURE A.1

Reductio ad Absurdum

It is a matter of logic that a statement cannot be both true and false. This fact is the basis for a method of proof called "reductio ad absurdum" or, more commonly, "proof by contradiction." The idea is to make the assumption that the conclusion of a statement is false and show that this leads to a contradiction of some sort. The underlying logic is that if $H \Rightarrow C$ is a true statement, then the statement

$$(H \text{ and } {\sim}C) \Rightarrow C$$

must be false, for otherwise C would be both true and false.

Sets

Many of the proofs in this text are concerned with ***sets*** (or collections) of objects, the objects being called the ***elements*** of the set. Although a set can generally include any kinds of objects, in linear algebra the objects are typically "scalars," "matrices," or "vectors" (terms that are all defined in the text). We assume that you are already familiar with the basic terminology and notation of sets, but we will review it quickly here.

Sets are generally denoted by capital letters and their elements by lowercase letters. One way to describe a set is to simply list its elements enclosed by braces; for example,

$$S = \{1, 3, 5\} \tag{13}$$

By agreement, *the elements of a set must all be different, and the order in which the elements are listed does not matter*. Thus, for example, the above set might also be written as

$$S = \{3, 5, 1\} \quad \text{or} \quad S = \{5, 1, 3\}$$

To indicate that an element a is a member of a set S we write $a \in S$ (read, "a belongs to S"), and to indicate that a is not a member of S we write $a \notin S$ (read, "a does not belong to S"). Thus, for the set in (13) we have

$$3 \in S \quad \text{and} \quad 4 \notin S$$

There are two common ways of denoting sets with infinitely many elements: If the elements have some obvious notational pattern, then the set can be denoted by explicitly specifying some initial elements and using dots to indicate that the remaining elements follow the same pattern. For example, the set of positive integers might be denoted as

$$S = \{1, 2, 3, \ldots\} \tag{14}$$

An alternative method for denoting the set S in (14) is to write

$$S = \{x \colon x \text{ is a positive integer}\}$$

where the right side is read, "the set of all x such that x is a positive integer." This is called ***set-builder*** notation. In general, set-builder notation has the form

$$S = \{x: \underline{\hspace{2cm}}\} \tag{15}$$

where the blank line is replaced by a description that defines those and only those elements in the set S. Of particular interest in this text are the set of real numbers, denoted by R; the set of points in the plane, denoted by R^2; and the set of points in three-dimensional space, denoted by R^3. The latter two can be described in set-builder notation as

$$R^2 = \{(x, y): x, y \in R\} \quad \text{and} \quad R^3 = \{(x, y, z): x, y, z \in R\}$$

Operations on Sets

If A and B are arbitrary sets, then the ***union*** of A and B, denoted by $A \cup B$, is the set of elements that belong to A or B or both; and the ***intersection*** of A and B, denoted by $A \cap B$, is the set of elements that belong to both A and B. These operations are illustrated in Figure A.2 using ***Venn diagrams***, named for the British logician John A. Venn (1834–1923). In those diagrams the sets A and B are the regions enclosed by the circles, and the sets $A \cup B$ and $A \cap B$ are shaded. In the event that the sets A and B have no common elements, then we say that the sets are ***disjoint*** and we write $A \cap B = \varnothing$, where the symbol $\varnothing$ denotes a set with no elements called the ***empty set***.

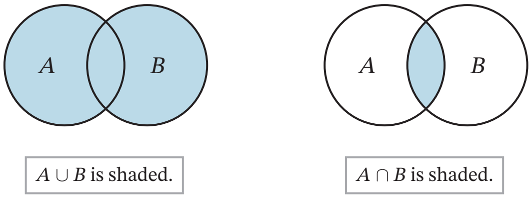


FIGURE A.2

If every element of a set A belongs as well to a set B, then we say that ***A is a subset of B*** and we write $A \subset B$. If $A \subset B$ and $B \subset A$, then A and B have exactly the same elements, so we say that A and B are ***equal*** and we write $A = B$.

Ordered Sets

In certain linear algebra problems the order in which elements are listed is important, so we will want to consider ***ordered sets***, that is, sets in which duplicate elements are not allowed but order matters. Thus, for example,

$$S_1 = \{3, 5, 1\} \quad \text{and} \quad S_2 = \{5, 1, 3\}$$

are the same *sets*, but not the same *ordered sets*.

How to Do Proofs

- A good first step in a proof is to write down in complete sentences what is given (i.e., the hypothesis H) and what is to be proved (i.e., the conclusion C).
- Once you clearly understand what is given and what is to be proved, you must decide whether you want to prove the theorem directly, or in contrapositive form, or by reductio ad absurdum. You might restate the theorem in the three ways and see which form seems most promising.
- Next, you might want to review earlier theorems that could be relevant to your proof.
- From this point on it is a matter of experience and intuition, but keep in mind that proving theorems is not an easy task, so don't be discouraged. As you read through the proofs in the text, observe the techniques and try to make them part of your own repertoire.
- Additional ideas on proving theorems can be found in the supplement to this text entitled "*How to Read and Do Proofs*" by Daniel Solow (see Preface for details).

Appendix B

Complex Numbers

Complex numbers arise naturally in the course of solving polynomial equations. For example, the solutions of the quadratic equation $ax^2 + bx + c = 0$, which are given by the quadratic formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

are complex numbers if the expression inside the radical is negative. In this appendix we will review some of the basic ideas about complex numbers that are used in this text.

Complex Numbers

To deal with the problem that the equation $x^2 = -1$ has no real solutions, mathematicians of the eighteenth century invented the "imaginary" number

$$i = \sqrt{-1}$$

which is assumed to have the property

$$i^2 = (\sqrt{-1})^2 = -1$$

but which otherwise has the algebraic properties of a real number. An expression of the form

$$a + bi \quad \text{or} \quad a + ib$$

in which a and b are *real* numbers is called a ***complex number***. Sometimes it will be convenient to use a single letter, typically z, to denote a complex number, in which case we write

$$z = a + bi \quad \text{or} \quad z = a + ib$$

The number a is called the ***real part*** of z and is denoted by $\text{Re}(z)$, and the number b is called the ***imaginary part*** of z and is denoted by $\text{Im}(z)$. Thus,

$$\begin{array}{ll}
\text{Re}(3 + 2i) = 3, & \text{Im}(3 + 2i) = 2 \\
\text{Re}(1 - 5i) = 1, & \text{Im}(1 - 5i) = \text{Im}(1 + (-5)i) = -5 \\
\text{Re}(7i) = \text{Re}(0 + 7i) = 0, & \text{Im}(7i) = \text{Im}(0 + 7i) = 7 \\
\text{Re}(4) = \text{Re}(4 + 0i) = 4, & \text{Im}(4) = \text{Im}(4 + 0i) = 0
\end{array}$$

Two complex numbers are considered ***equal*** if and only if their real parts are equal and their imaginary parts are equal; that is,

$$a + bi = c + di \quad \text{if and only if} \quad a = c \text{ and } b = d$$

A complex number $z = bi$ whose real part is zero is said to be ***pure imaginary***. A complex number $z = a$ whose imaginary part is zero is a real number, so the real numbers can be viewed as a subset of the complex numbers.

Complex numbers are added, subtracted, and multiplied in accordance with the standard rules of algebra but with $i^2 = -1$:

$$\boxed{(a + bi) + (c + di) = (a + c) + (b + d)i} \tag{1}$$

$$\boxed{(a + bi) - (c + di) = (a - c) + (b - d)i} \tag{2}$$

$$\boxed{(a + bi)(c + di) = (ac - bd) + (ad + bc)i} \tag{3}$$

Multiplication formula (3) is obtained by expanding the left side and using the fact that $i^2 = -1$. Also note that if $b = 0$, then the multiplication formula simplifies to

$$\boxed{a(c + di) = ac + adi} \tag{4}$$

The set of complex numbers with these operations is commonly denoted by the symbol C and is called the ***complex number system***.

EXAMPLE 1 | Multiplying Complex Numbers

As a practical matter, it is usually more convenient to compute products of complex numbers by expansion, rather than substituting in (3). For example,

$$(3 - 2i)(4 + 5i) = 12 + 15i - 8i - 10i^2 = (12 + 10) + 7i = 22 + 7i$$

The Complex Plane

A complex number $z = a + bi$ can be associated with the ordered pair (a, b) of real numbers and represented geometrically by a point or a vector in the xy-plane (**Figure B.1**). We call this the ***complex plane***. Points on the x-axis have an imaginary part of zero and hence correspond to real numbers, whereas points on the y-axis have a real part of zero and correspond to pure imaginary numbers. Accordingly, we call the x-axis the ***real axis*** and the y-axis the ***imaginary axis*** (**Figure B.2**).

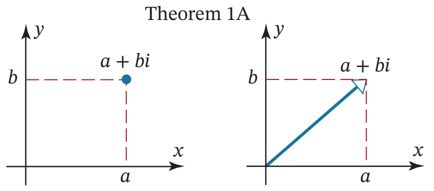


FIGURE B.1

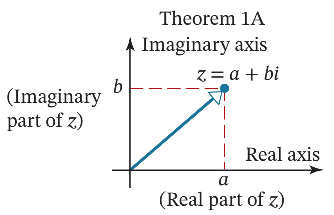


FIGURE B.2

Complex numbers can be added, subtracted, or multiplied by real numbers geometrically by performing these operations on their associated vectors (**Figure B.3**, for example). In this sense the complex number system C is closely related to R^2, the main difference being that complex numbers can be multiplied to produce other complex numbers, whereas there is no multiplication operation on R^2 that produces other vectors in R^2 (the dot product defined in Section 3.2 produces a scalar, not a vector in R^2).

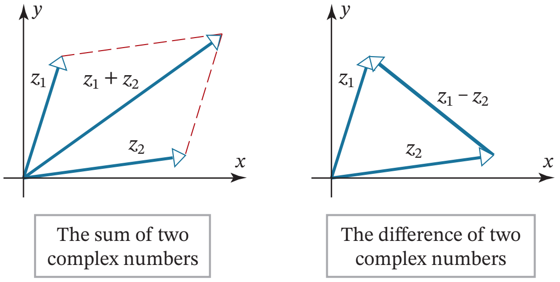


FIGURE B.3

If $z = a + bi$ is a complex number, then the ***complex conjugate*** of z, or more simply, the ***conjugate*** of z, is denoted by $\bar{z}$ (read, "z bar") and is defined by

$$\boxed{\bar{z} = a - bi} \tag{5}$$

Numerically, $\bar{z}$ is obtained from z by reversing the sign of the imaginary part, and geometrically it is obtained by reflecting the vector for z about the real axis (**Figure B.4**).

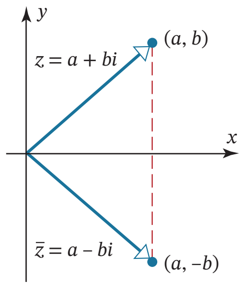


FIGURE B.4

EXAMPLE 2 | Some Complex Conjugates

$$\begin{array}{ll} z = 3 + 4i & \bar{z} = 3 - 4i \\ z = -2 - 5i & \bar{z} = -2 + 5i \\ z = i & \bar{z} = -i \\ z = 7 & \bar{z} = 7 \end{array}$$

Remark The last computation in this example illustrates the fact that a real number is equal to its complex conjugate. More generally, $z = \bar{z}$ if and only if z is a real number.

The following computation shows that the product of a complex number $z = a + bi$ and its conjugate $z = a - bi$ is a nonnegative real number:

$$z\bar{z} = (a + bi)(a - bi) = a^2 - abi + bai - b^2i^2 = a^2 + b^2 \tag{6}$$

You will recognize that

$$\sqrt{z\bar{z}} = \sqrt{a^2 + b^2}$$

is the length of the vector corresponding to z (**Figure B.5**); we call this length the ***modulus*** (or ***absolute value*** of z) and denote it by $|z|$. Thus,

$$\boxed{|z| = \sqrt{z\bar{z}} = \sqrt{a^2 + b^2}} \tag{7}$$

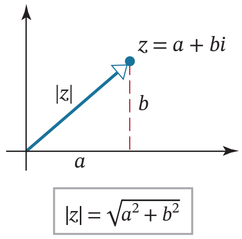


FIGURE B.5

Note that if $b = 0$, then $z = a$ is a real number and $|z| = \sqrt{a^2} = |a|$, which tells us that the modulus of a real number is the same as its absolute value.

EXAMPLE 3 | Some Modulus Computations

$$\begin{array}{ll} z = 3 + 4i & |z| = \sqrt{3^2 + 4^2} = 5 \\ z = -4 - 5i & |z| = \sqrt{(-4)^2 + (-5)^2} = \sqrt{41} \\ z = i & |z| = \sqrt{0^2 + 1^2} = 1 \end{array}$$

Reciprocals and Division

If $z \neq 0$, then the ***reciprocal*** (or ***multiplicative inverse***) of z is denoted by $1/z$ (or z^{-1}) and is defined by the property

$$\left(\frac{1}{z}\right)z = 1$$

This equation has a unique solution for $1/z$, which we can obtain by multiplying both sides by $\bar{z}$ and using the fact that $z\bar{z} = |z|^2$ [see (7)]. This yields

$$\boxed{\frac{1}{z} = \frac{\bar{z}}{|z|^2}} \tag{8}$$

If $z_2 \neq 0$, then the ***quotient*** z_1/z_2 is defined to be the product of z_1 and $1/z_2$. This yields the formula

$$\boxed{\frac{z_1}{z_2} = \frac{\bar{z}_2}{|z_2|^2} z_1 = \frac{z_1 \bar{z}_2}{|z_2|^2}} \tag{9}$$

Observe that the expression on the right side of (9) results if the numerator and denominator of z_1/z_2 are multiplied by $\bar{z}_2$. As a practical matter, this is often the best way to perform divisions of complex numbers.

EXAMPLE 4 | Division of Complex Numbers

Let $z_1 = 3 + 4i$ and $z_2 = 1 - 2i$. Express z_1/z_2 in the form $a + bi$.

Solution We will multiply the numerator and denominator of z_1/z_2 by $\bar{z}_2 = 1 + 2i$. This yields

$$\begin{aligned} \frac{z_1}{z_2} = \frac{z_1 \bar{z}_2}{z_2 \bar{z}_2} &= \frac{3 + 4i}{1 - 2i} \cdot \frac{1 + 2i}{1 + 2i} \\ &= \frac{3 + 6i + 4i + 8i^2}{1 - 4i^2} \\ &= \frac{-5 + 10i}{5} \\ &= -1 + 2i \end{aligned}$$

The following theorems list some useful properties of the modulus and conjugate operations.

Theorem B1

The following results hold for any complex numbers z, z_1, and z_2.

(*a*) $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$

(*b*) $\overline{z_1 - z_2} = \bar{z}_1 - \bar{z}_2$

(*c*) $\overline{z_1 z_2} = \bar{z}_1 \bar{z}_2$

(*d*) $\overline{z_1/z_2} = \bar{z}_1/\bar{z}_2$

(*e*) $\bar{\bar{z}} = z$

Theorem B2

The following results hold for any complex numbers z, z_1, and z_2.

(*a*) $|\bar{z}| = |z|$

(*b*) $|z_1 z_2| = |z_1||z_2|$

(*c*) $|z_1/z_2| = |z_1|/|z_2|$

(*d*) $|z_1 + z_2| \leq |z_1| + |z_2|$

Polar Form of a Complex Number

If $z = a + bi$ is a nonzero complex number, and if ϕ is an angle from the real axis to the vector z, then, as suggested in **Figure B.6**, the real and imaginary parts of z can be expressed as

$$a = |z|\cos\phi \quad \text{and} \quad b = |z|\sin\phi \tag{10}$$

Thus, the complex number $z = a + bi$ can be expressed as

$$\boxed{z = |z|(\cos\phi + i\sin\phi)} \tag{11}$$

which is called a ***polar form*** of z. The angle ϕ in this formula is called an ***argument*** of z. The argument of z is not unique because we can add or subtract any multiple of 2π to it to obtain a different argument of z. However, there is only one argument whose radian measure satisfies

$$-\pi < \phi \le \pi \tag{12}$$

This is called the ***principal argument*** of z.

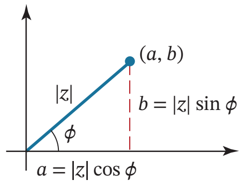


FIGURE B.6

EXAMPLE 5 | Polar Form of a Complex Number

Express $z = 1 - \sqrt{3}i$ in polar form using the principal argument.

Solution The modulus of z is

$$|z| = \sqrt{1^2 + (-\sqrt{3})^2} = \sqrt{4} = 2$$

Thus, it follows from (10) with $a = 1$ and $b = -\sqrt{3}$ that

$$1 = 2\cos\phi \quad \text{and} \quad -\sqrt{3} = 2\sin\phi$$

and this implies that

$$\cos\phi = \frac{1}{2} \quad \text{and} \quad \sin\phi = -\frac{\sqrt{3}}{2}$$

The unique angle ϕ that satisfies these equations and whose radian measure satisfies (12) is $\phi = -\pi/3$ (**Figure B.7**). Thus, a polar form of z is

$$z = 2\left(\cos\left(-\tfrac{\pi}{3}\right) + i\sin\left(-\tfrac{\pi}{3}\right)\right) = 2\left(\cos\tfrac{\pi}{3} - i\sin\tfrac{\pi}{3}\right)$$

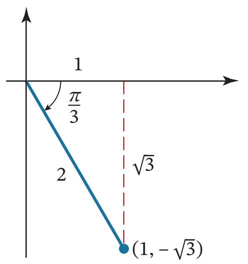


FIGURE B.7

Geometric Interpretation of Multiplication and Division of Complex Numbers

We will now show how polar forms of complex numbers provide geometric interpretations of multiplication and division. Let

$$z_1 = |z_1|(\cos\phi_1 + i\sin\phi_1) \quad \text{and} \quad z_2 = |z_2|(\cos\phi_2 + i\sin\phi_2)$$

be polar forms of the nonzero complex numbers z_1 and z_2. Multiplying, we obtain

$$z_1z_2 = |z_1||z_2|[(\cos\phi_1\cos\phi_2 - \sin\phi_1\sin\phi_2) + i(\sin\phi_1\cos\phi_2 + \cos\phi_1\sin\phi_2)]$$

Now applying the trigonometric identities

$$\begin{aligned}\cos(\phi_1 + \phi_2) &= \cos\phi_1\cos\phi_2 - \sin\phi_1\sin\phi_2 \\ \sin(\phi_1 + \phi_2) &= \sin\phi_1\cos\phi_2 + \cos\phi_1\sin\phi_2\end{aligned}$$

yields

$$\boxed{z_1z_2 = |z_1||z_2|[\cos(\phi_1 + \phi_2) + i\sin(\phi_1 + \phi_2)]} \tag{13}$$

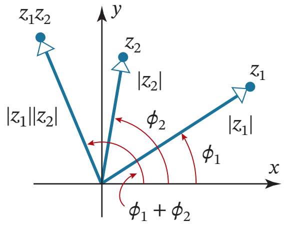


FIGURE B.8

which is a polar form of the complex number that has modulus $|z_1||z_2|$ and argument $\phi_1+\phi_2$. Thus, we have shown that *multiplying two complex numbers has the geometric effect of multiplying their moduli and adding their arguments* (**Figure B.8**).

Similar kinds of computations show that

$$\boxed{\frac{z_1}{z_2}=\frac{|z_1|}{|z_2|}[\cos(\phi_1-\phi_2)+i\sin(\phi_1-\phi_2)]} \tag{14}$$

which tells us that *dividing complex numbers has the geometric effect of dividing their moduli and subtracting their arguments* (each in the appropriate order).

EXAMPLE 6 | Multiplying and Dividing in Polar Form

Use polar forms of the complex numbers $z_1=1+\sqrt{3}\,i$ and $z_2=\sqrt{3}+i$ to compute z_1z_2 and z_1/z_2.

Solution Polar forms of these complex numbers are

$$z_1=2\left(\cos\frac{\pi}{3}+i\sin\frac{\pi}{3}\right)\quad\text{and}\quad z_2=2\left(\cos\frac{\pi}{6}+i\sin\frac{\pi}{6}\right)$$

(verify). Thus, it follows from (13) that

$$z_1z_2=4\left[\cos\left(\frac{\pi}{3}+\frac{\pi}{6}\right)+i\sin\left(\frac{\pi}{3}+\frac{\pi}{6}\right)\right]=4\left[\cos\left(\frac{\pi}{2}\right)+i\sin\left(\frac{\pi}{2}\right)\right]=4i$$

and from (14) that

$$\frac{z_1}{z_2}=1\cdot\left[\cos\left(\frac{\pi}{3}-\frac{\pi}{6}\right)+i\sin\left(\frac{\pi}{3}-\frac{\pi}{6}\right)\right]=\cos\left(\frac{\pi}{6}\right)+i\sin\left(\frac{\pi}{6}\right)=\frac{\sqrt{3}}{2}+\frac{1}{2}i$$

As a check, let us calculate z_1z_2 and z_1/z_2 directly:

$$z_1z_2=(1+\sqrt{3}\,i)(\sqrt{3}+i)=\sqrt{3}+i+3i+\sqrt{3}\,i^2=4i$$

$$\frac{z_1}{z_2}=\frac{1+\sqrt{3}\,i}{\sqrt{3}+i}=\frac{1+\sqrt{3}\,i}{\sqrt{3}+i}\cdot\frac{\sqrt{3}-i}{\sqrt{3}-i}=\frac{\sqrt{3}-i+3i-\sqrt{3}\,i^2}{3-i^2}=\frac{2\sqrt{3}+2i}{4}=\frac{\sqrt{3}}{2}+\frac{1}{2}i$$

which agrees with the results obtained using polar forms.

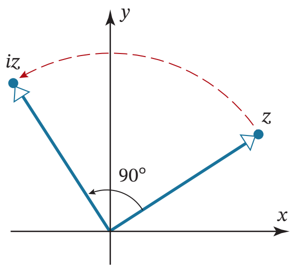


FIGURE B.9

Remark The complex number i has a modulus of 1 and a principal argument of $\pi/2$. Thus, if z is a complex number, then iz has the same modulus as z but its argument is greater by $\pi/2$ ($=90°$); that is, multiplication by i has the geometric effect of rotating the vector z counterclockwise by 90° (**Figure B.9**).

DeMoivre's Formula

If n is a positive integer, and if z is a nonzero complex number with polar form

$$z=|z|(\cos\phi+i\sin\phi)$$

then raising z to the nth power yields

$$z^n=\underbrace{z\cdot z\cdot\cdots\cdot z}_{n\text{ factors}}=|z|^n[\cos(\underbrace{\phi+\phi+\cdots+\phi}_{n\text{ terms}})]+i[\sin(\underbrace{\phi+\phi+\cdots+\phi}_{n\text{ terms}})]$$

which we can write more succinctly as

$$\boxed{z^n=|z|^n(\cos n\phi+i\sin n\phi)} \tag{15}$$

In the special case where $|z| = 1$ this formula simplifies to

$$z^n = \cos n\phi + i \sin n\phi$$

which, using the polar form for z, becomes

$$\boxed{(\cos\phi + i\sin\phi)^n = \cos n\phi + i\sin n\phi} \tag{16}$$

This result is called ***DeMoivre's formula***, named for the French mathematician Abraham de Moivre (1667–1754).

Euler's Formula

If θ is a real number, say the radian measure of some angle, then the ***complex exponential*** function $e^{i\theta}$ is defined to be

$$\boxed{e^{i\theta} = \cos\theta + i\sin\theta} \tag{17}$$

which is sometimes called ***Euler's formula***, named for the Swiss mathematician Leonhard Euler (1707–1783). One motivation for this formula comes from the Maclaurin series in calculus. Readers who have studied infinite series in calculus can deduce (17) by formally substituting $i\theta$ for x in the Maclaurin series for e^x and writing

$$\begin{aligned} e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \frac{(i\theta)^6}{6!} + \cdots \\ &= 1 + i\theta - \frac{\theta^2}{2!} - i\frac{\theta^3}{3!} + \frac{\theta^4}{4!} + i\frac{\theta^5}{5!} - \frac{\theta^6}{6!} + \cdots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right) \\ &= \cos\theta + i\sin\theta \end{aligned}$$

where the last step follows from the Maclaurin series for $\cos\theta$ and $\sin\theta$.

If $z = a + bi$ is any complex number, then the ***complex exponential*** e^z is defined to be

$$e^z = e^{a+bi} = e^a e^{ib} = e^a(\cos b + i\sin b) \tag{18}$$

It can be proved that complex exponentials satisfy the standard laws of exponents. Thus, for example,

$$e^{z_1}e^{z_2} = e^{z_1+z_2}, \quad \frac{e^{z_1}}{e^{z_2}} = e^{z_1-z_2}, \quad \frac{1}{e^z} = e^{-z}$$

Answers to Exercises

Chapter 1

Exercise Set 1.1 (page 8)

1. (a), (c), and (f) are linear equations; (b), (d), and (e) are not linear equations.

3. **a.** $\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2 \end{aligned}$ **b.** $\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned}$ **c.** $\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 &= b_2 \end{aligned}$

5. **a.** $\begin{aligned} 2x_1 \qquad\quad &= 0 \\ 3x_1 - 4x_2 &= 0 \\ x_2 &= 1 \end{aligned}$ **b.** $\begin{aligned} 3x_1 \qquad\quad - 2x_3 &= 5 \\ 7x_1 + x_2 + 4x_3 &= -3 \\ - 2x_2 + x_3 &= 7 \end{aligned}$

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

9. (a), (d), and (e) are solutions; (b) and (c) are not solutions.

11. **a.** No points of intersection

b. Infinitely many points of intersection: $x = \frac{1}{2} + 2t,\ y = t$

c. One point of intersection: $(-8, -4)$

13. **a.** $x = \frac{3}{7} + \frac{5}{7}t,\ y = t$

b. $x_1 = \frac{7}{3} + \frac{5}{3}r - \frac{4}{3}s,\ x_2 = r,\ x_3 = s$

c. $x_1 = -\frac{1}{8} + \frac{1}{4}r - \frac{5}{8}s + \frac{3}{4}t,\ x_2 = r,\ x_3 = s,\ x_4 = t$

d. $v = \frac{8}{3}t_1 - \frac{2}{3}t_2 + \frac{1}{3}t_3 - \frac{4}{3}t_4,\ w = t_1,\ x = t_2,\ y = t_3,\ z = t_4$

15. **a.** $x = \frac{1}{2} + \frac{3}{2}t,\ y = t$

b. $x_1 = -4 - 3r + s,\ x_2 = r,\ x_3 = s$

17. **a.** Add 2 times the second row to the first row.

b. Add the third row to the first row, or interchange the first row and the third row.

19. **a.** All values of $k \neq 2$

b. All values of k

25. $\begin{aligned} 2x + 3y + \ z &= 7 \\ 2x + \ y + 3z &= 9 \\ 4x + 2y + 5z &= 16 \end{aligned}$

27. $\begin{aligned} x + y + \ z &= 12 \\ 2x + y + 2z &= 5 \\ -x \qquad + \ z &= 1 \end{aligned}$

True/False 1.1

a. True **b.** False **c.** True **d.** True **e.** False **f.** False **g.** True **h.** False

Exercise Set 1.2 (page 22)

1. **a.** Both **b.** Both **c.** Both **d.** Both **e.** Both **f.** Both **g.** Row echelon form

3. **a.** Rows 1, 2, and 3 are the pivot rows; columns 1, 2, and 3 are the pivot columns.
$x = -37,\ y = -8,\ z = 5$

b. Rows 1, 2, and 3 are the pivot rows; columns 1, 2, and 3 are the pivot columns.
$w = -10 + 13t,\ x = -5 + 13t,\ y = 2 - t,\ z = t$

c. Rows 1, 2, and 3 are the pivot rows; columns 1, 3, and 4 are the pivot columns.
$x_1 = -11 - 7s + 2t,\ x_2 = s,\ x_3 = -4 - 3t,\ x_4 = 9 - 3t,\ x_5 = t$

d. Rows 1 and 2 are the pivot rows; columns 1 and 2 are the pivot columns.
No solution

5. $x_1 = 3,\ x_2 = 1,\ x_3 = 2$ **7.** $x = -1 + t,\ y = 2s,\ z = s,\ w = t$ **9.** $x_1 = 3,\ x_2 = 1,\ x_3 = 2$

11. $x = -1 + t,\ y = 2s,\ z = s,\ w = t$ **13.** Has nontrivial solutions **15.** $x_1 = 0,\ x_2 = 0,\ x_3 = 0$

17. $x_1 = -\frac{1}{4}s,\ x_2 = -\frac{1}{4}s - t,\ x_3 = s,\ x_4 = t$ **19.** $w = t,\ x = -t,\ y = t,\ z = 0$ **21.** $I_1 = -1,\ I_2 = 0,\ I_3 = 1,\ I_4 = 2$

23. **a.** Consistent; unique solution

b. Consistent; infinitely many solutions

c. Inconsistent

d. Insufficient information provided

25. No solutions when $a = -4$; infinitely many solutions when $a = 4$; one solution for all values $a \neq -4$ and $a \neq 4$

27. $-a + b + c = 0$

29. $x = \frac{2}{3}a - \frac{1}{9}b,\ y = -\frac{1}{3}a + \frac{2}{9}b$

31. E.g., $\begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (other answers are possible)

35. $x = \pm 1,\ y = \pm\sqrt{3},\ z = \pm\sqrt{2}$

37. $a = 1, b = -6, c = 2, d = 10$

39. The nonhomogeneous system has only one solution.

True/False 1.2

a. True **b.** False **c.** False **d.** True **e.** True **f.** False **g.** True **h.** False **i.** False

Exercise Set 1.3 (page 37)

1. **a.** Undefined **b.** Defined; 4×4 matrix **c.** Defined; 4×2 matrix
d. Defined; 5×2 matrix **e.** Defined; 4×5 matrix **f.** Defined; 5×5 matrix

3. **a.** $\begin{bmatrix} 7 & 6 & 5 \\ -2 & 1 & 3 \\ 7 & 3 & 7 \end{bmatrix}$ **b.** $\begin{bmatrix} -5 & 4 & -1 \\ 0 & -1 & -1 \\ -1 & 1 & 1 \end{bmatrix}$ **c.** $\begin{bmatrix} 15 & 0 \\ -5 & 10 \\ 5 & 5 \end{bmatrix}$ **d.** $\begin{bmatrix} -7 & -28 & -14 \\ -21 & -7 & -35 \end{bmatrix}$

e. Undefined **f.** $\begin{bmatrix} 22 & -6 & 8 \\ -2 & 4 & 6 \\ 10 & 0 & 4 \end{bmatrix}$ **g.** $\begin{bmatrix} -39 & -21 & -24 \\ 9 & -6 & -15 \\ -33 & -12 & -30 \end{bmatrix}$ **h.** $\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$

i. 5 **j.** -25 **k.** 168 **l.** Undefined

5. **a.** $\begin{bmatrix} 12 & -3 \\ -4 & 5 \\ 4 & 1 \end{bmatrix}$ **b.** Undefined **c.** $\begin{bmatrix} 42 & 108 & 75 \\ 12 & -3 & 21 \\ 36 & 78 & 63 \end{bmatrix}$ **d.** $\begin{bmatrix} 3 & 45 & 9 \\ 11 & -11 & 17 \\ 7 & 17 & 13 \end{bmatrix}$ **e.** $\begin{bmatrix} 3 & 45 & 9 \\ 11 & -11 & 17 \\ 7 & 17 & 13 \end{bmatrix}$

f. $\begin{bmatrix} 21 & 17 \\ 17 & 35 \end{bmatrix}$ **g.** $\begin{bmatrix} 0 & -2 & 11 \\ 12 & 1 & 8 \end{bmatrix}$ **h.** $\begin{bmatrix} 12 & 6 & 9 \\ 48 & -20 & 14 \\ 24 & 8 & 16 \end{bmatrix}$ **i.** 61 **j.** 35 **k.** 28 **l.** 99

7. **a.** $\begin{bmatrix} 67 & 41 & 41 \end{bmatrix}$ **b.** $\begin{bmatrix} 63 & 67 & 57 \end{bmatrix}$ **c.** $\begin{bmatrix} 41 \\ 21 \\ 67 \end{bmatrix}$ **d.** $\begin{bmatrix} 6 \\ 6 \\ 63 \end{bmatrix}$ **e.** $\begin{bmatrix} 24 & 56 & 97 \end{bmatrix}$ **f.** $\begin{bmatrix} 76 \\ 98 \\ 97 \end{bmatrix}$

9. **a.** first column of $AA = 3\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} + 6\begin{bmatrix} -2 \\ 5 \\ 4 \end{bmatrix} + 0\begin{bmatrix} 7 \\ 4 \\ 9 \end{bmatrix}$

second column of $AA = -2\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} + 5\begin{bmatrix} -2 \\ 5 \\ 4 \end{bmatrix} + 4\begin{bmatrix} 7 \\ 4 \\ 9 \end{bmatrix}$

third column of $AA = 7\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} + 4\begin{bmatrix} -2 \\ 5 \\ 4 \end{bmatrix} + 9\begin{bmatrix} 7 \\ 4 \\ 9 \end{bmatrix}$

b. first column of $BB = 6\begin{bmatrix} 6 \\ 0 \\ 7 \end{bmatrix} + 0\begin{bmatrix} -2 \\ 1 \\ 7 \end{bmatrix} + 7\begin{bmatrix} 4 \\ 3 \\ 5 \end{bmatrix}$

second column of $BB = -2\begin{bmatrix} 6 \\ 0 \\ 7 \end{bmatrix} + 1\begin{bmatrix} -2 \\ 1 \\ 7 \end{bmatrix} + 7\begin{bmatrix} 4 \\ 3 \\ 5 \end{bmatrix}$

third column of $BB = 4\begin{bmatrix} 6 \\ 0 \\ 7 \end{bmatrix} + 3\begin{bmatrix} -2 \\ 1 \\ 7 \end{bmatrix} + 5\begin{bmatrix} 4 \\ 3 \\ 5 \end{bmatrix}$

11. **a.** $A = \begin{bmatrix} 2 & -3 & 5 \\ 9 & -1 & 1 \\ 1 & 5 & 4 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 7 \\ -1 \\ 0 \end{bmatrix}; \begin{bmatrix} 2 & -3 & 5 \\ 9 & -1 & 1 \\ 1 & 5 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 7 \\ -1 \\ 0 \end{bmatrix}$

b. $A = \begin{bmatrix} 4 & 0 & -3 & 1 \\ 5 & 1 & 0 & -8 \\ 2 & -5 & 9 & -1 \\ 0 & 3 & -1 & 7 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 2 \end{bmatrix}; \begin{bmatrix} 4 & 0 & -3 & 1 \\ 5 & 1 & 0 & -8 \\ 2 & -5 & 9 & -1 \\ 0 & 3 & -1 & 7 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 2 \end{bmatrix}$

13. **a.**
$$\begin{aligned} 5x_1 + 6x_2 - 7x_3 &= 2 \\ -x_1 - 2x_2 + 3x_3 &= 0 \\ 4x_2 - x_3 &= 3 \end{aligned}$$

b.
$$\begin{aligned} x + y + z &= 2 \\ 2x + 3y &= 2 \\ 5x - 3y - 6z &= -9 \end{aligned}$$

15. $k = -1$

17. $\begin{bmatrix}4\\2\end{bmatrix}\begin{bmatrix}0 & 1 & 2\end{bmatrix}+\begin{bmatrix}-3\\-1\end{bmatrix}\begin{bmatrix}-2 & 3 & 1\end{bmatrix}=\begin{bmatrix}0 & 4 & 8\\0 & 2 & 4\end{bmatrix}+\begin{bmatrix}6 & -9 & -3\\2 & -3 & -1\end{bmatrix}=\begin{bmatrix}6 & -5 & 5\\2 & -1 & 3\end{bmatrix}$

19. $\begin{bmatrix}1\\4\end{bmatrix}\begin{bmatrix}1 & 2\end{bmatrix}+\begin{bmatrix}2\\5\end{bmatrix}\begin{bmatrix}3 & 4\end{bmatrix}+\begin{bmatrix}3\\6\end{bmatrix}\begin{bmatrix}5 & 6\end{bmatrix}=\begin{bmatrix}1 & 2\\4 & 8\end{bmatrix}+\begin{bmatrix}6 & 8\\15 & 20\end{bmatrix}+\begin{bmatrix}15 & 18\\30 & 36\end{bmatrix}=\begin{bmatrix}22 & 28\\49 & 64\end{bmatrix}$

21. $\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\\x_6\end{bmatrix}=\begin{bmatrix}0\\0\\0\\0\\0\\\frac{1}{3}\end{bmatrix}+r\begin{bmatrix}-3\\1\\0\\0\\0\\0\end{bmatrix}+s\begin{bmatrix}-4\\0\\-2\\1\\0\\0\end{bmatrix}+t\begin{bmatrix}-2\\0\\0\\0\\1\\0\end{bmatrix}$

23. $a=4, b=-6, c=-1, d=1$

27. The only matrix satisfying the given condition is $A=\begin{bmatrix}1 & 1 & 0\\1 & -1 & 0\\0 & 0 & 0\end{bmatrix}$.

29. **a.** $\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$ and $\begin{bmatrix}-1 & -1\\-1 & -1\end{bmatrix}$

b. Four square roots can be found: $\begin{bmatrix}\sqrt{5} & 0\\0 & 3\end{bmatrix}$, $\begin{bmatrix}-\sqrt{5} & 0\\0 & 3\end{bmatrix}$, $\begin{bmatrix}\sqrt{5} & 0\\0 & -3\end{bmatrix}$, and $\begin{bmatrix}-\sqrt{5} & 0\\0 & -3\end{bmatrix}$.

33. The matrix product represents $\begin{bmatrix}\text{the total cost of items purchased in January}\\\text{the total cost of items purchased in February}\\\text{the total cost of items purchased in March}\\\text{the total cost of items purchased in April}\end{bmatrix}$.

True/False 1.3

a. True **b.** False **c.** False **d.** False **e.** True **f.** False **g.** False **h.** True **i.** True **j.** True
k. True **l.** False **m.** True **n.** True **o.** False

Exercise Set 1.4 (page 51)

5. $\begin{bmatrix}\frac{1}{5} & \frac{3}{20}\\-\frac{1}{5} & \frac{1}{10}\end{bmatrix}$ **7.** $\begin{bmatrix}\frac{1}{2} & 0\\0 & \frac{1}{3}\end{bmatrix}$ **9.** $\begin{bmatrix}\frac{1}{2}(e^x+e^{-x}) & -\frac{1}{2}(e^x-e^{-x})\\-\frac{1}{2}(e^x-e^{-x}) & \frac{1}{2}(e^x+e^{-x})\end{bmatrix}$ **15.** $\begin{bmatrix}\frac{2}{7} & 1\\\frac{1}{7} & \frac{3}{7}\end{bmatrix}$ **17.** $\begin{bmatrix}-\frac{9}{13} & \frac{1}{13}\\\frac{2}{13} & -\frac{6}{13}\end{bmatrix}$

19. **a.** $\begin{bmatrix}41 & 15\\30 & 11\end{bmatrix}$ **b.** $\begin{bmatrix}11 & -15\\-30 & 41\end{bmatrix}$ **c.** $\begin{bmatrix}6 & 2\\4 & 2\end{bmatrix}$ **21.** **a.** $\begin{bmatrix}1 & 1\\2 & -1\end{bmatrix}$ **b.** $\begin{bmatrix}20 & 7\\14 & 6\end{bmatrix}$ **c.** $\begin{bmatrix}36 & 13\\26 & 10\end{bmatrix}$

23. The matrices commute if $c=0$ and $a=d$. **25.** $x_1=\frac{1}{23}, x_2=\frac{13}{23}$ **27.** $x_1=-\frac{1}{11}, x_2=\frac{6}{11}$

31. **a.** E.g., $A=\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$ and $B=\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}$

b. $(A+B)(A-B)=A(A-B)+B(A-B)=A^2-AB+BA-B^2$

c. $AB=BA$

35. No **37.** Invertible; $A^{-1}=\begin{bmatrix}\frac{1}{2} & \frac{1}{2} & -\frac{1}{2}\\-\frac{1}{2} & \frac{1}{2} & \frac{1}{2}\\\frac{1}{2} & -\frac{1}{2} & \frac{1}{2}\end{bmatrix}$ **39.** B^{-1}

True/False 1.4

a. False **b.** False **c.** False **d.** False **e.** False **f.** True **g.** True **h.** True **i.** False **j.** True **k.** False

Exercise Set 1.5 (page 60)

1. **a.** Elementary **b.** Not elementary **c.** Not elementary **d.** Not elementary

3. **a.** Add 3 times the second row to the first row: $\begin{bmatrix}1 & 3\\0 & 1\end{bmatrix}$ **b.** Multiply the first row by $-\frac{1}{7}$: $\begin{bmatrix}-\frac{1}{7} & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}$

c. Add 5 times the first row to the third row: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 5 & 0 & 1 \end{bmatrix}$ **d.** Interchange the first and third rows: $\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

5. **a.** Interchange the first and second rows: $EA = \begin{bmatrix} 3 & -6 & -6 & -6 \\ -1 & -2 & 5 & -1 \end{bmatrix}$

b. Add -3 times the second row to the third row: $EA = \begin{bmatrix} 2 & -1 & 0 & -4 & -4 \\ 1 & -3 & -1 & 5 & 3 \\ -1 & 9 & 4 & -12 & -10 \end{bmatrix}$

c. Add 4 times the third row to the first row: $EA = \begin{bmatrix} 13 & 28 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

7. **a.** $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ **b.** $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ **c.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}$ **d.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$ **9.** **a.** $\begin{bmatrix} -7 & 4 \\ 2 & -1 \end{bmatrix}$ **b.** Not invertible

11. **a.** The inverse is $\begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix}$. **b.** Not invertible **13.** $\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$ **15.** $\begin{bmatrix} \frac{7}{2} & 0 & -3 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$

17. $\begin{bmatrix} \frac{1}{4} & \frac{1}{2} & -3 & 0 \\ -\frac{1}{8} & \frac{1}{4} & -\frac{3}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{40} & -\frac{1}{20} & -\frac{1}{10} & -\frac{1}{5} \end{bmatrix}$ **19.** **a.** $\begin{bmatrix} \frac{1}{k_1} & 0 & 0 & 0 \\ 0 & \frac{1}{k_2} & 0 & 0 \\ 0 & 0 & \frac{1}{k_3} & 0 \\ 0 & 0 & 0 & \frac{1}{k_4} \end{bmatrix}$ **b.** $\begin{bmatrix} \frac{1}{k} & -\frac{1}{k} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{k} & -\frac{1}{k} \\ 0 & 0 & 0 & 1 \end{bmatrix}$

21. Any value of c other than 0 and 1

23. $A = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -8 \end{bmatrix}\begin{bmatrix} 1 & 5 \\ 0 & 1 \end{bmatrix}$; $A^{-1} = \begin{bmatrix} 1 & -5 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{8} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$

(Answer is not unique.)

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

(Answer is not unique.)

27. Add -1 times the first row to the second row; add -1 times the second row to the first row; add -1 times the first row to the third row. (Answer is not unique.)

True/False 1.5

a. False **b.** True **c.** True **d.** True **e.** True **f.** True **g.** False

Exercise Set 1.6 (page 67)

1. $x_1 = 3, x_2 = -1$ **3.** $x_1 = -1, x_2 = 4, x_3 = -7$ **5.** $x = 1, y = 5$, and $z = -1$ **7.** $x_1 = 2b_1 - 5b_2, x_2 = -b_1 + 3b_2$

9. **i.** $x_1 = \frac{22}{17}, x_2 = \frac{1}{17}$ **ii.** $x_1 = \frac{21}{17}, x_2 = \frac{11}{17}$

11. **i.** $x_1 = \frac{7}{15}, x_2 = \frac{4}{15}$ **ii.** $x_1 = \frac{34}{15}, x_2 = \frac{28}{15}$ **iii.** $x_1 = \frac{19}{15}, x_2 = \frac{13}{15}$ **iv.** $x_1 = -\frac{1}{5}, x_2 = \frac{3}{5}$

13. The system is consistent for all values of b_1 and b_2. **15.** $b_1 = b_2 + b_3$ **17.** $b_1 = b_3 + b_4$ and $b_2 = 2b_3 + b_4$

19. $X = \begin{bmatrix} 11 & 12 & -3 & 27 & 26 \\ -6 & -8 & 1 & -18 & -17 \\ -15 & -21 & 9 & -38 & -35 \end{bmatrix}$

True/False 1.6

a. True **b.** True **c.** True **d.** True **e.** True **f.** True **g.** True

Exercise Set 1.7 (page 74)

1. **a.** Upper triangular and invertible
b. Lower triangular and not invertible
c. Diagonal, upper triangular, lower triangular, and invertible
d. Upper triangular and not invertible

3. $\begin{bmatrix} 6 & 3 \\ 4 & -1 \\ 4 & 10 \end{bmatrix}$ **5.** $\begin{bmatrix} -15 & 10 & 0 & 20 & -20 \\ 2 & -10 & 6 & 0 & 6 \\ 18 & -6 & -6 & -6 & -6 \end{bmatrix}$ **7.** $A^2 = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}, A^{-2} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{4} \end{bmatrix}, A^{-k} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{(-2)^k} \end{bmatrix}$

9. $A^2 = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{9} & 0 \\ 0 & 0 & \frac{1}{16} \end{bmatrix}, A^{-2} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 16 \end{bmatrix}, A^{-k} = \begin{bmatrix} 2^k & 0 & 0 \\ 0 & 3^k & 0 \\ 0 & 0 & 4^k \end{bmatrix}$ **11.** $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ **13.** $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

15. **a.** $\begin{bmatrix} au & av \\ bw & bx \\ cy & cz \end{bmatrix}$ **b.** $\begin{bmatrix} ra & sb & tc \\ ua & vb & wc \\ xa & yb & zc \end{bmatrix}$ **17.** **a.** $\begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 3 & 7 & 2 \\ 3 & 1 & -8 & -3 \\ 7 & -8 & 0 & 9 \\ 2 & -3 & 9 & 0 \end{bmatrix}$

19. Not invertible **21.** Invertible **23.** $-3, 5, -6$ **25.** $a = -8$ **27.** All x such that $x \neq 1$, $x \neq -2$, and $x \neq 4$

29. They are reciprocals of the corresponding diagonal entries of the matrix A. **31.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$

37. **a.** Symmetric **b.** Not symmetric (unless $n = 1$) **c.** Symmetric **d.** Not symmetric (unless $n = 1$)

39. $\begin{bmatrix} 1 & 10 \\ 0 & -2 \end{bmatrix}$ **41.** **a.** $\begin{bmatrix} 0 & 0 & 4 \\ 0 & 0 & 1 \\ -4 & -1 & 0 \end{bmatrix}$ **b.** $\begin{bmatrix} 0 & 0 & -8 \\ 0 & 0 & -4 \\ 8 & 4 & 0 \end{bmatrix}$ **43.** No

True/False 1.7

a. True **b.** False **c.** False **d.** True **e.** True **f.** False **g.** False **h.** True **i.** True **j.** False **k.** False
l. False **m.** True

Exercise Set 1.8 (page 88)

1. **a.** Domain: R^2; codomain: R^3
b. Domain: R^3; codomain: R^2
c. Domain: R^3; codomain: R^3
d. Domain: R^6; codomain: R

3. **a.** Domain: R^2; codomain: R^2
b. Domain: R^2; codomain: R^3

5. **a.** Domain: R^3; codomain: R^2
b. Domain: R^2; codomain: R^3

7. **a.** Domain: R^2; codomain: R^2
b. Domain: R^3; codomain: R^2

9. Domain: R^2; codomain: R^3

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

13. **a.** $\begin{bmatrix} 0 & 1 \\ -1 & 0 \\ 1 & 3 \\ 1 & -1 \end{bmatrix}$ **b.** $\begin{bmatrix} 7 & 2 & -1 & 1 \\ 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}$ **c.** $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ **d.** $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \end{bmatrix}$

15. $\begin{bmatrix} 3 & 5 & -1 \\ 4 & -1 & 1 \\ 3 & 2 & -1 \end{bmatrix}$; $T(-1, 2, 4) = (3, -2, -3)$

17. **a.** $\begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix}$; $T(-1, 4) = (5, 4)$ **b.** $\begin{bmatrix} 2 & -1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$; $T(2, 1, -3) = (0, -2, 0)$

19. **a.** $T_A(\mathbf{x}) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ **b.** $T_A(\mathbf{x}) = \begin{bmatrix} -1 & 2 & 0 \\ 3 & 1 & 5 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 13 \end{bmatrix}$ **25.** No, unless $b = 0$

27. $\begin{bmatrix} 1 & 0 & 4 \\ 3 & 0 & -3 \\ 0 & 1 & -1 \end{bmatrix}$; $T(\mathbf{x}) = \begin{bmatrix} 2 \\ 6 \\ 1 \end{bmatrix}$

29. **a.** $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$ **b.** $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ **c.** $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

31. **a.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} 2 \\ -5 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ -5 \\ -3 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ -5 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix}$ **c.** $\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ -5 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -5 \\ 3 \end{bmatrix}$

33. **a.** $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ -5 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ **b.** $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ -5 \end{bmatrix} = \begin{bmatrix} 0 \\ -5 \end{bmatrix}$

35. **a.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 3 \end{bmatrix}$ **c.** $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$

37. **a.** $\begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}\begin{bmatrix} 3 \\ -4 \end{bmatrix} = \begin{bmatrix} \frac{3\sqrt{3}}{2} + 2 \\ \frac{3}{2} - 2\sqrt{3} \end{bmatrix} \approx \begin{bmatrix} 4.60 \\ -1.96 \end{bmatrix}$ **b.** $\begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 3 \\ -4 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} - 2\sqrt{3} \\ -\frac{3\sqrt{3}}{2} - 2 \end{bmatrix} \approx \begin{bmatrix} -1.96 \\ -4.60 \end{bmatrix}$

c. $\begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}\begin{bmatrix} 3 \\ -4 \end{bmatrix} = \begin{bmatrix} \frac{7\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix} \approx \begin{bmatrix} 4.95 \\ -0.71 \end{bmatrix}$ **d.** $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 3 \\ -4 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$

39. $(a + c, b + d)$

41. **a.** $T_A(\mathbf{e}_1) = \begin{bmatrix} -1 \\ 2 \\ 4 \end{bmatrix}, T_A(\mathbf{e}_2) = \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}, T_A(\mathbf{e}_3) = \begin{bmatrix} 0 \\ 2 \\ -3 \end{bmatrix}$ **b.** $\begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}$ **c.** $\begin{bmatrix} 0 \\ 14 \\ -21 \end{bmatrix}$

43. Reflection about the xy-plane: $T(1, 2, 3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$;

Reflection about the xz-plane: $T(1, 2, 3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$;

Reflection about the yz-plane: $T(1, 2, 3) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}$

45. $A = \begin{bmatrix} 5 & -4 \\ -11 & 9 \end{bmatrix}$

49. Rotation through the angle 2θ

True/False 1.8

a. False **b.** False **c.** True **d.** False **e.** True **f.** False **g.** False

Exercise Set 1.9 (page 96)

1. **a.** Operators do not commute. **b.** Operators do not commute.

3. The operators commute.

5. The standard matrices for $T_B \circ T_A$ and $T_A \circ T_B$ are $\begin{bmatrix} -10 & -7 \\ 5 & -10 \end{bmatrix}$ and $\begin{bmatrix} -8 & -3 \\ 13 & -12 \end{bmatrix}$, respectively.

7. **a.** $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ **b.** $\begin{bmatrix} 0 & -\frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} \end{bmatrix}$ **c.** $\begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}$

9. **a.** $\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ **c.** $\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

11. a. The standard matrices for T_1 and T_2 are $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and $\begin{bmatrix} 3 & 0 \\ 2 & 4 \end{bmatrix}$, respectively.

b. The standard matrices for $T_2 \circ T_1$ and $T_1 \circ T_2$ are $\begin{bmatrix} 3 & 3 \\ 6 & -2 \end{bmatrix}$ and $\begin{bmatrix} 5 & 4 \\ 1 & -4 \end{bmatrix}$, respectively.

c. $T_1(T_2(x_1, x_2)) = (5x_1 + 4x_2, x_1 - 4x_2)$; $T_2(T_1(x_1, x_2)) = (3x_1 + 3x_2, 6x_1 - 2x_2)$

13. a. The standard matrices for T_1 and T_2 are

$\begin{bmatrix} 1 & -1 \\ -1 & 2 \\ 3 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 4 & 0 \\ 1 & 2 & 0 \end{bmatrix}$, respectively.

b. The standard matrices for $T_2 \circ T_1$ and $T_1 \circ T_2$ are

$\begin{bmatrix} -4 & 8 \\ -1 & 3 \end{bmatrix}$ and $\begin{bmatrix} -1 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 12 & 0 \end{bmatrix}$, respectively.

c. $T_1(T_2(x_1, x_2, x_3)) = (-x_1 - 2x_2, 2x_1, 12x_2)$;

$T_2(T_1(x_1, x_2)) = (-4x_1 + 8x_2, -x_1 + 3x_2)$

15. a. The standard matrices for T_1 and T_2 are

$\begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$, respectively.

b. The standard matrix for $T_2 \circ T_1$ is $\begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 2 & 0 \end{bmatrix}$.

c. The domain of T_1 does not equal the codomain of T_2.

d. $(T_2 \circ T_1)(x, y) = (x, 2x - y, 2x)$

17. a. $\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 8 & 4 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$; the operator is not invertible.

b. $\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} -1 & 3 & 2 \\ 2 & 0 & 4 \\ 1 & 3 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$; the operator is not invertible.

19. a. Invertible; standard matrix of T^{-1}: $\begin{bmatrix} \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix}$; $T^{-1}(w_1, w_2) = (\frac{1}{3}w_1 - \frac{2}{3}w_2, \frac{1}{3}w_1 + \frac{1}{3}w_2)$

b. Not invertible

21. a. Invertible; reflection about the x-axis in R^2

b. Invertible; 300° rotation about the origin in R^2

c. Not invertible

23. a. Invertible; $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$

b. Not invertible

25. a. Reflection about the line $y = x$ followed by reflection about the origin

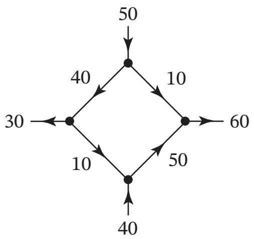
b. $T_A = T_B \circ T_C$, where $B = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ and $C = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

True/False 1.9

a. False **b.** True **c.** True **d.** False **e.** True **f.** False **g.** True

Exercise Set 1.10 (page 108)

1.



3. a.
$$\begin{aligned} x_2 - x_3 \qquad &= 100 \\ x_3 - x_4 &= -500 \\ x_1 - x_2 \qquad\qquad &= 300 \\ -x_1 \qquad\qquad + x_4 &= 100 \end{aligned}$$

b. $x_1 = -100 + s, x_2 = -400 + s, x_3 = -500 + s, x_4 = s$

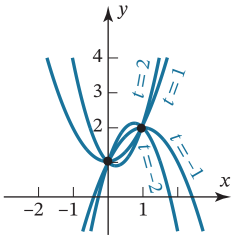
c. To keep the traffic flowing on all roads, the flow from A to B must exceed 500 vehicles per hour.

5. $I_1 = 2.6\text{A}, I_2 = -0.4\text{A}, I_3 = 2.2\text{A}$ **7.** $I_1 = I_4 = I_5 = I_6 = 0.5\text{A}, I_2 = I_3 = 0\text{A}$ **9.** $C_3H_8 + 5O_2 \rightarrow 3CO_2 + 4H_2O$

11. $CH_3COF + H_2O \rightarrow CH_3COOH + HF$ **13.** $2 - 2x + x^2$ **15.** $1 + \frac{13}{6}x - \frac{1}{6}x^3$

17. a. $p(x) = 1 + (1 - t)x + tx^2$

b.



True/False 1.10

a. False **b.** False **c.** True **d.** False **e.** False

Exercise Set 1.11 (page 114)

1. a. $\begin{bmatrix} 0.50 & 0.25 \\ 0.25 & 0.10 \end{bmatrix}$

b. M must produce approximately \$25,290.32 worth of mechanical work and B must produce approximately \$22,580.65 worth of body work.

3. a. $\begin{bmatrix} 0.10 & 0.60 & 0.40 \\ 0.30 & 0.20 & 0.30 \\ 0.40 & 0.10 & 0.20 \end{bmatrix}$ **b.** $\begin{bmatrix} \$31{,}500 \\ \$26{,}500 \\ \$26{,}300 \end{bmatrix}$ **5.** $\mathbf{x} \approx \begin{bmatrix} 123.08 \\ 202.56 \end{bmatrix}$

True/False 1.11

a. False **b.** True **c.** False **d.** True **e.** True

Chapter 1 Supplementary Exercises (page 115)

1.
$$\begin{aligned} 3x_1 - x_2 \qquad\quad + 4x_4 &= 1 \\ 2x_1 \qquad + 3x_3 + 3x_4 &= -1 \end{aligned}$$
$x_1 = -\frac{3}{2}s - \frac{3}{2}t - \frac{1}{2}, x_2 = -\frac{9}{2}s - \frac{1}{2}t - \frac{5}{2}, x_3 = s, x_4 = t$

3.
$$\begin{aligned} 2x_1 - 4x_2 + \ x_3 &= 6 \\ -4x_1 \qquad\quad + 3x_3 &= -1 \\ x_2 - \ x_3 &= 3 \end{aligned}$$
$x_1 = -\frac{17}{2}, x_2 = -\frac{26}{3}, x_3 = -\frac{35}{3}$

5. $x' = \frac{3}{5}x + \frac{4}{5}y, y' = -\frac{4}{5}x + \frac{3}{5}y$ **7.** $x = 4, y = 2, z = 3$

9. a. $a \neq 0$ and $b \neq 2$ **b.** $a \neq 0$ and $b = 2$ **c.** $a = 0$ and $b = 2$ **d.** $a = 0$ and $b \neq 2$

11. $\begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix}$ **13. a.** $\begin{bmatrix} -1 & 3 & -1 \\ 6 & 0 & 1 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & -2 \\ 3 & 1 \end{bmatrix}$ **c.** $\begin{bmatrix} -\frac{113}{37} & -\frac{160}{37} \\ -\frac{20}{37} & -\frac{46}{37} \end{bmatrix}$ **15.** $a = 1, b = -2, c = 3$

Chapter 2

Exercise Set 2.1 (page 124)

1. $M_{11} = 29, C_{11} = 29$
$M_{12} = 21, C_{12} = -21$
$M_{13} = 27, C_{13} = 27$
$M_{21} = -11, C_{21} = 11$
$M_{22} = 13, C_{22} = 13$
$M_{23} = -5, C_{23} = 5$
$M_{31} = -19, C_{31} = -19$
$M_{32} = -19, C_{32} = 19$
$M_{33} = 19, C_{33} = 19$

3. **a.** $M_{13} = 0, C_{13} = 0$
b. $M_{23} = -96, C_{23} = 96$
c. $M_{22} = -48, C_{22} = -48$
d. $M_{21} = 72, C_{21} = -72$

5. 22; $\begin{bmatrix} \frac{2}{11} & \frac{-5}{22} \\ \frac{1}{11} & \frac{3}{22} \end{bmatrix}$ **7.** 59; $\begin{bmatrix} \frac{-2}{59} & \frac{-7}{59} \\ \frac{7}{59} & \frac{-5}{59} \end{bmatrix}$

9. $a^2 - 5a + 21$ **11.** -65 **13.** -123 **15.** $\lambda = -3$ or $\lambda = 1$ **17.** $\lambda = 1$ or $\lambda = -1$ **19.** (all parts) -123 **21.** -40

23. 0 **25.** -240 **27.** -1 **29.** 0 **31.** 6 **33.** **a.** The determinant is 1. **b.** The determinant is 1.

35. $d_1 + \lambda = d_2$ **37.** If $n = 1$ then the determinant is 1. If $n \geq 2$ then the determinant is 0.

True/False 2.1

a. False **b.** False **c.** True **d.** True **e.** True **f.** True **g.** False **h.** False **i.** False **j.** True

Exercise Set 2.2 (page 131)

5. -5 **7.** -1 **9.** 33 **11.** 6 **13.** -2 **15.** -6 **17.** 72 **19.** -6

21. 18 **31.** -24 **33.** $\det(B) = (-1)^{\lfloor n/2 \rfloor} \det(A)$

True/False 2.2

a. True **b.** True **c.** False **d.** False **e.** True **f.** True

Exercise Set 2.3 (page 142)

5. $\det(A + B) \neq \det(A) + \det(B)$ **7.** Invertible **9.** Invertible **11.** Not invertible **13.** Invertible

15. $k \neq \frac{5-\sqrt{17}}{2}$ and $k \neq \frac{5+\sqrt{17}}{2}$ **17.** $k \neq -1$ **19.** Invertible; $A^{-1} = \begin{bmatrix} 3 & -5 & -5 \\ -3 & 4 & 5 \\ 2 & -2 & -3 \end{bmatrix}$ **21.** Invertible; $A^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{3}{2} & 1 \\ 0 & 1 & \frac{3}{2} \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$

23. Invertible; $A^{-1} = \begin{bmatrix} -4 & 3 & 0 & -1 \\ 2 & -1 & 0 & 0 \\ -7 & 0 & -1 & 8 \\ 6 & 0 & 1 & -7 \end{bmatrix}$ **25.** $x = \frac{3}{11}, y = \frac{2}{11}, z = -\frac{1}{11}$ **27.** $x_1 = -\frac{30}{11}, x_2 = -\frac{38}{11}, x_3 = -\frac{40}{11}$

29. Cramer's rule does not apply. **31.** $y = 0$ **33.** **a.** -189 **b.** $-\frac{1}{7}$ **c.** $-\frac{8}{7}$ **d.** $-\frac{1}{56}$ **e.** 7

35. **a.** 189 **b.** $\frac{1}{7}$ **c.** $\frac{8}{7}$ **d.** $\frac{1}{56}$

True/False 2.3

a. False **b.** False **c.** True **d.** False **e.** True **f.** True **g.** True **h.** True **i.** True **j.** True **k.** True
l. False

Chapter 2 Supplementary Exercises (page 144)

1. -18 **3.** 24 **5.** -10 **7.** 329 **9.** Exercise 3: 24; Exercise 4: 0; Exercise 5: -10; Exercise 6: -48

11. The matrices in Exercises 1–3 are invertible; the matrix in Exercise 4 is not. **13.** $-b^2+5b-21$ **15.** -120

17. $\begin{bmatrix} -\frac{1}{6} & \frac{1}{9} \\ \frac{1}{6} & \frac{2}{9} \end{bmatrix}$ **19.** $\begin{bmatrix} \frac{1}{8} & -\frac{1}{8} & -\frac{3}{8} \\ \frac{1}{8} & \frac{5}{24} & -\frac{1}{24} \\ \frac{1}{4} & -\frac{7}{12} & -\frac{1}{12} \end{bmatrix}$ **21.** $\begin{bmatrix} \frac{1}{5} & \frac{2}{5} & -\frac{1}{10} \\ \frac{1}{5} & -\frac{3}{5} & \frac{2}{5} \\ -\frac{2}{5} & \frac{6}{5} & -\frac{3}{10} \end{bmatrix}$ **23.** $\begin{bmatrix} \frac{10}{329} & -\frac{2}{329} & \frac{52}{329} & -\frac{27}{329} \\ \frac{55}{329} & -\frac{11}{329} & -\frac{43}{329} & \frac{16}{329} \\ -\frac{3}{47} & \frac{10}{47} & -\frac{25}{47} & -\frac{6}{47} \\ -\frac{31}{329} & \frac{72}{329} & \frac{102}{329} & -\frac{15}{329} \end{bmatrix}$

25. $x' = \frac{3}{5}x + \frac{4}{5}y, y' = \frac{3}{5}y - \frac{4}{5}x$ **29. b.** $\cos\beta = \frac{a^2+c^2-b^2}{2ac}, \cos\gamma = \frac{a^2+b^2-c^2}{2ab}$

Chapter 3

Exercise Set 3.1 (page 156)

1. a. $(3,-4)$ **b.** $(2,-3,4)$ **3. a.** $(-1,3)$ **b.** $(-3,6,1)$ **5. a.** $(2,3)$ **b.** $(-2,-2,-1)$

7. a. $(-1,2,-4)$ is one possible answer. **b.** $(7,-2,-6)$ is one possible answer.

9. a. $(1,-4)$ **b.** $(-12,8)$ **c.** $(38,28)$ **d.** $(4,29)$

11. a. $(-1,9,-11,1)$ **b.** $(-13,13,-36,-2)$ **c.** $(-90,-114,60,-36)$ **d.** $(27,29,-27,9)$

13. $\left(-\frac{25}{3},7,-\frac{32}{3},-\frac{2}{3}\right)$ **15. a.** Not parallel to $\mathbf{u}$ **b.** Parallel to $\mathbf{u}$ **c.** Parallel to $\mathbf{u}$ **17.** $a=3, b=-1$

19. $c_1=2, c_2=-1, c_3=5$ **23. a.** $\left(\frac{9}{2},-\frac{1}{2},-\frac{1}{2}\right)$ **b.** $\left(\frac{23}{4},-\frac{9}{4},\frac{1}{4}\right)$ **25. a.** $(-2,5)$ **b.** $(3,-8)$ **27.** $(7,-3,-19)$

29. a. $\mathbf{0}$ **b.** $\mathbf{0}$ **c.** $-\mathbf{a}$ **31.** Magnitude of $\mathbf{F}$ is $\sqrt{84}$ lb ≈ 9.17 lb; the angle with the positive x-axis $\approx -70.9°$

33. $\frac{500}{1+\sqrt{3}}$ lb ≈ 183.01 lb and $\frac{750\sqrt{2}}{3+\sqrt{3}}$ lb ≈ 224.14 lb

True/False 3.1

a. False **b.** False **c.** False **d.** True **e.** True **f.** False **g.** False **h.** True **i.** False **j.** True **k.** False

Exercise Set 3.2 (page 170)

1. a. $\|\mathbf{v}\| = 2\sqrt{3}$; $\frac{1}{\|\mathbf{v}\|}\mathbf{v} = \left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)$; $-\frac{1}{\|\mathbf{v}\|}\mathbf{v} = \left(-\frac{1}{\sqrt{3}},-\frac{1}{\sqrt{3}},-\frac{1}{\sqrt{3}}\right)$

b. $\|\mathbf{v}\| = \sqrt{15}$; $\frac{1}{\|\mathbf{v}\|}\mathbf{v} = \left(\frac{1}{\sqrt{15}},0,\frac{2}{\sqrt{15}},\frac{1}{\sqrt{15}},\frac{3}{\sqrt{15}}\right)$; $-\frac{1}{\|\mathbf{v}\|}\mathbf{v} = \left(-\frac{1}{\sqrt{15}},0,-\frac{2}{\sqrt{15}},-\frac{1}{\sqrt{15}},-\frac{3}{\sqrt{15}}\right)$

3. a. $\sqrt{83}$ **b.** $\sqrt{17}+\sqrt{26}$ **c.** $2\sqrt{3}$ **d.** $\sqrt{466}$ **5. a.** $\sqrt{2570}$ **b.** $3\sqrt{46}-10\sqrt{21}+\sqrt{42}$ **c.** $2\sqrt{966}$

7. $k=\frac{5}{7}$ or $k=-\frac{5}{7}$ **9. a.** $\mathbf{u}\cdot\mathbf{v}=-8$; $\mathbf{u}\cdot\mathbf{u}=26$; $\mathbf{v}\cdot\mathbf{v}=24$ **b.** $\mathbf{u}\cdot\mathbf{v}=0$; $\mathbf{u}\cdot\mathbf{u}=54$; $\mathbf{v}\cdot\mathbf{v}=21$

11. a. $d(\mathbf{u},\mathbf{v})=\sqrt{14}$; $\cos\theta = \frac{5}{\sqrt{51}}$; the angle is acute. **13.** $\frac{45\sqrt{3}}{2}$

b. $d(\mathbf{u},\mathbf{v})=\sqrt{59}$; $\cos\theta = \frac{-4}{\sqrt{6}\sqrt{45}}$; the angle is obtuse.

15. a. Does not make sense; $\mathbf{v}\cdot\mathbf{w}$ is a scalar, whereas the dot product is only defined for vectors.
b. Makes sense
c. Does not make sense; $\mathbf{u}\cdot\mathbf{v}$ is a scalar, whereas the norm is only defined for vectors.
d. Makes sense

25. $71°, 61°, 36°$

True/False 3.2

a. True **b.** True **c.** False **d.** True **e.** True **f.** False **g.** False **h.** False **i.** True **j.** True

Exercise Set 3.3 (page 181)

1. a. Orthogonal **b.** Not orthogonal **c.** Not orthogonal **d.** Not orthogonal

3. $-2(x+1)+(y-3)-(z+2)=0$ **5.** $2z=0$ **7.** Not parallel **9.** Parallel **11.** Not perpendicular

13. a. $\frac{2}{5}$ **b.** $\frac{18}{\sqrt{22}}$ **15.** $(0,0),(6,2)$ **17.** $\left(-\frac{16}{13},0,-\frac{80}{13}\right),\left(\frac{55}{13},1,-\frac{11}{13}\right)$ **19.** $\left(\frac{1}{5},-\frac{1}{5},\frac{1}{10},-\frac{1}{10}\right),\left(\frac{9}{5},\frac{6}{5},\frac{9}{10},\frac{21}{10}\right)$

21. 1 **23.** $\frac{1}{\sqrt{17}}$ **25.** $\frac{5}{3}$ **27.** $\frac{11}{\sqrt{6}}$ **29.** $\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},-\frac{1}{\sqrt{3}}\right)$ is one possible answer. **31.** Yes

35. The standard matrix: $\begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$; $H_{\pi/3}(3,4) = \left(-\frac{3}{2} + 2\sqrt{3}, \frac{3\sqrt{3}}{2} + 2\right) \approx (1.96, 4.60)$

37. The standard matrix: $\begin{bmatrix} \frac{1}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{3}{4} \end{bmatrix}$; $P_{\pi/3}(3,4) = \left(\frac{3}{4} + \sqrt{3}, \frac{3\sqrt{3}}{4} + 3\right) \approx (2.48, 4.30)$

41. $\frac{50{,}000}{\sqrt{2}}$ Nm $\approx$ 35,355 Nm

True/False 3.3

a. True **b.** True **c.** True **d.** True **e.** True **f.** False **g.** False

Exercise Set 3.4 (page 189)

1. Vector equation: $(x, y) = (-4, 1) + t(0, -8)$;
parametric equations: $x = -4, y = 1 - 8t$

3. Vector equation: $(x, y, z) = t(-3, 0, 1)$;
parametric equations: $x = -3t, y = 0, z = t$

5. Point: $(3, -6)$; vector: $(-5, -1)$

7. Point: $(4, 6)$; vector: $(-6, -6)$

9. Vector equation: $(x, y, z) = (-3, 1, 0) + t_1(0, -3, 6) + t_2(-5, 1, 2)$;
parametric equations: $x = -3 - 5t_2, y = 1 - 3t_1 + t_2, z = 6t_1 + 2t_2$

11. Vector equation: $(x, y, z) = (-1, 1, 4) + t_1(6, -1, 0) + t_2(-1, 3, 1)$;
parametric equations: $x = -1 + 6t_1 - t_2, y = 1 - t_1 + 3t_2, z = 4 + t_2$

13. Vector equation: $(x, y) = t(3, 2)$;
parametric equations: $x = 3t$ and $y = 2t$

15. Vector equation: $(x, y, z) = t_1(5, 0, 4) + t_2(0, 1, 0)$;
parametric equations: $x = 5t_1, y = t_2$, and $z = 4t_1$

17. $x_1 = -s - t, x_2 = s, x_3 = t$

19. $x_1 = \frac{3}{7}r - \frac{19}{7}s - \frac{8}{7}t, x_2 = -\frac{2}{7}r + \frac{1}{7}s + \frac{3}{7}t, x_3 = r, x_4 = s, x_5 = t$

21. a. $\begin{aligned} x + y + z &= 0 \\ -2x + 3y \quad &= 0 \end{aligned}$ **b.** A straight line passing through the origin **c.** $x = -\frac{3}{5}t, y = -\frac{2}{5}t, z = t$

23. b. $\mathbf{x} = (5, -9) + t(3, 10); x = 5 + 3t, \ y = -9 + 10t$

True/False 3.4

a. True **b.** False **c.** True **d.** True **e.** True

Exercise Set 3.5 (page 198)

1. a. $(32, -6, -4)$ **b.** $(-32, 6, 4)$ **c.** $(52, -29, 10)$ **d.** 0 **e.** $(0, 0, 0)$ **f.** $(0, 0, 0)$

3. $\|\mathbf{u} \times \mathbf{w}\|^2 = 1125$ **5.** $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (-14, -20, -82)$ **7.** $\mathbf{u} \times \mathbf{v} = (18, 36, -18)$ **9.** $\sqrt{59}$ **11.** 3

13. 7 **15.** $\frac{\sqrt{374}}{2}$ **17.** 16 **19.** The vectors do not lie in the same plane. **21.** -92 **23.** abc **25. a.** -3 **b.** 3 **c.** 3

27. a. $\frac{\sqrt{26}}{2}$ **b.** $\frac{\sqrt{26}}{3}$ **29.** $2(\mathbf{v} \times \mathbf{u})$ **31. a.** $1500\sqrt{2}$ Nm $\approx$ 2121.32 Nm **b.** 132°, 109°, 132° **39. a.** $\frac{17}{6}$ **b.** $\frac{1}{2}$

True/False 3.5

a. True **b.** True **c.** False **d.** True **e.** False **f.** False

Chapter 3 Supplementary Exercises (page 200)

1. a. $(13, -3, 10)$ **b.** $\sqrt{70}$ **c.** $3\sqrt{86}$ **d.** $\left(-\frac{8}{9}, \frac{20}{9}, \frac{20}{9}\right)$ **e.** -122 **f.** $(-3150, -2430, 1170)$

3. a. $(-5, -12, 20, -2)$ **b.** $\sqrt{106}$ **c.** $\sqrt{2810}$ **d.** $\left(-\frac{135}{77}, -\frac{15}{77}, \frac{90}{77}, \frac{90}{77}\right)$

5. The plane containing A, B, and C **7.** $(-1, -1, 5)$ **9.** $\sqrt{\frac{14}{17}}$ **11.** $\frac{11}{\sqrt{35}}$

13. Vector equation: $(x, y, z) = (-2, 1, 3) + t_1(1, -2, -2) + t_2(5, -1, -5)$;
parametric equations: $x = -2 + t_1 + 5t_2, y = 1 - 2t_1 - t_2, z = 3 - 2t_1 - 5t_2$

15. Vector equation: $(x, y) = (0, -3) + t(8, -1)$;
parametric equations: $x = 8t, y = -3 - t$

17. Vector equation: $(x, y) = (0, -5) + t(1, 3)$;
parametric equations: $x = t, y = -5 + 3t$

19. $3(x+1) + 6(y-5) + 2(z-6) = 0$ **21.** $-18(x-9) - 51y - 24(z-4) = 0$ **25.** A plane through the origin

Chapter 4

Exercise Set 4.1 (page 209)

1. a. $\mathbf{u} + \mathbf{v} = (2, 6)$; $k\mathbf{u} = (0, 6)$ **c.** Axioms 1–5 **3.** Vector space **5.** Not a vector space; Axioms 5 and 6 fail.

7. Not a vector space; Axiom 8 fails. **9.** Vector space **11.** Vector space **19.** $\frac{1}{u} = u^{-1}$

True/False 4.1

a. True **b.** False **c.** False **d.** False **e.** True **f.** False

Exercise Set 4.2 (page 218)

1. (a), (c) **3.** (a), (c), (d) **5.** (a), (b) **7.** (a) **9.** (a) **11.** (a) **13.** (b) **15.** (a) **17.** (a), (b), (c), (d)

19. a. Line; $x = -\frac{1}{2}t, y = -\frac{3}{2}t, z = t$ **b.** Origin **c.** Plane; $x - 3y + z = 0$ **d.** Line; $x = -3t, y = -2t, z = t$

True/False 4.2

a. True **b.** True **c.** False **d.** False **e.** False **f.** True **g.** False **h.** True

Exercise Set 4.3 (page 226)

1. (a), (c) **3.** (a), (b)

5. a. $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = -3A + 12B - 13C + 2D$ **b.** $\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} = A + B + C + D$

7. a. The vectors span R^3. **b.** The vectors do not span R^3. **9.** The polynomials do not span P_2.

11. a. The matrices do not span M_{22}. **b.** The matrices do not span M_{22}. **c.** The matrices span M_{22}.

13. a. $\mathbf{u}$ is not in the span. **b.** $\mathbf{u}$ is in the span.

15. a. The set spans W. **b.** The set spans W.

17. a. The set spans R^2. **b.** The set does not span R^2.

19. $\mathbf{p}_1 = 0\mathbf{q}_1 + \mathbf{q}_2$, $\mathbf{p}_2 = \frac{1}{2}\mathbf{q}_1 + \mathbf{q}_2$; $\mathbf{q}_1 = -2\mathbf{p}_1 + 2\mathbf{p}_2$, $\mathbf{q}_2 = \mathbf{p}_1 + 0\mathbf{p}_2$

21. $\mathbf{v} = (-21, -7)$, $\mathbf{w} = (24, 12)$

True/False 4.3

a. True **b.** False **c.** False **d.** True **e.** True **f.** False **g.** False

Exercise Set 4.4 (page 236)

1. a. $\mathbf{u}_2 = -5\mathbf{u}_1$ **b.** A set of 3 vectors in R^2 must be linearly dependent by Theorem 4.3.3.
c. $\mathbf{p}_2 = 2\mathbf{p}_1$ **d.** $A = (-1)B$ **3. a.** Linearly dependent **b.** Linearly independent

5. a. Linearly independent **b.** Linearly independent

7. a. The vectors do not lie in a plane. **b.** The vectors lie in a plane.

9. b. $\mathbf{v}_1 = \frac{2}{7}\mathbf{v}_2 - \frac{3}{7}\mathbf{v}_3$; $\mathbf{v}_2 = \frac{7}{2}\mathbf{v}_1 + \frac{3}{2}\mathbf{v}_3$; $\mathbf{v}_3 = -\frac{7}{3}\mathbf{v}_1 + \frac{2}{3}\mathbf{v}_2$ **11.** $\lambda = -\frac{1}{2}, \lambda = 1$

13. a. Linearly independent **b.** Linearly dependent **15. a.** Linearly independent **b.** Linearly dependent

True/False 4.4

a. False **b.** True **c.** False **d.** True **e.** True **f.** False **g.** True **h.** False

Exercise Set 4.5 (page 245)

11. a. $\left(\frac{5}{28}, \frac{3}{14}\right)$ **b.** $\left(a, \frac{b-a}{2}\right)$ **13. a.** $(3, -2, 1)$ **b.** $(-2, 0, 1)$ **15.** $A = 1A_1 - 1A_2 + 1A_3 - 1A_4$; $(A)_S = (1, -1, 1, -1)$

17. $\mathbf{p} = 7\mathbf{p}_1 - 8\mathbf{p}_2 + 3\mathbf{p}_3$; $(\mathbf{p})_S = (7, -8, 3)$ **21. a.** Linearly independent **b.** Linearly dependent

23. a. $(2, 0)$ **b.** $\left(\frac{2}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right)$ **c.** $(0, 1)$ **d.** $\left(\frac{2a}{\sqrt{3}}, b - \frac{a}{\sqrt{3}}\right)$ **25. b.** $(3, 4, 2, 1)$

27. a. $(20, 17, 2)$ **b.** $3x^2 + 8x - 1$ **c.** $\begin{bmatrix} -21 & -103 \\ -106 & 30 \end{bmatrix}$

True/False 4.5

a. False **b.** False **c.** True **d.** True **e.** False

Exercise Set 4.6 (page 254)

1. Basis: $\{(1, 0, 1)\}$; dimension: 1 **3.** No basis; dimension: 0 **5.** Basis: $\{(3, 1, 0), (-1, 0, 1)\}$; dimension: 2

7. a. Basis: $\left\{\left(\frac{2}{3}, 1, 0\right), \left(-\frac{5}{3}, 0, 1\right)\right\}$; dimension: 2 **b.** Basis: $\{(1, 1, 0), (0, 0, 1)\}$; dimension: 2

c. Basis: $\{(2, -1, 4)\}$; dimension: 1 **d.** Basis: $S = \{(1, 1, 0), (0, 1, 1)\}$; dimension: 2

9. a. n **b.** $\frac{n(n+1)}{2}$ **c.** $\frac{n(n+1)}{2}$

11. b. Dimension: 2 **c.** Basis: $\{-1 + x, -1 + x^2\}$ **13.** $\mathbf{e}_2$ and $\mathbf{e}_3$ (The answer is not unique.)

15. $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{e}_1$ form a basis for R^3 (The answer is not unique.) **17.** $\{\mathbf{v}_1, \mathbf{v}_2\}$ (The answer is not unique.)

19. a. 1 **b.** 2 **c.** 1

27. a. $\{-1 + x - 2x^2, 3 + 3x + 6x^2, 9\}$ (The answer is not unique.) **b.** $\{1 + x, x^2\}$ (The answer is not unique.)

c. $\{1 + x - 3x^2\}$ (The answer is not unique.)

True/False 4.6

a. True **b.** True **c.** False **d.** True **e.** True **f.** True **g.** True **h.** True **i.** True **j.** False **k.** False

Exercise Set 4.7 (page 261)

1. a. $\begin{bmatrix} \frac{13}{10} & -\frac{1}{2} \\ -\frac{2}{5} & 0 \end{bmatrix}$ **b.** $\begin{bmatrix} 0 & -\frac{5}{2} \\ -2 & -\frac{13}{2} \end{bmatrix}$ **c.** $[\mathbf{w}]_B = \begin{bmatrix} -\frac{17}{10} \\ \frac{8}{5} \end{bmatrix}$; $[\mathbf{w}]_{B'} = \begin{bmatrix} -4 \\ -7 \end{bmatrix}$

3. a. $\begin{bmatrix} 3 & 2 & \frac{5}{2} \\ -2 & -3 & -\frac{1}{2} \\ 5 & 1 & 6 \end{bmatrix}$ **b.** $[\mathbf{w}]_B = \begin{bmatrix} 9 \\ -9 \\ -5 \end{bmatrix}$; $[\mathbf{w}]_{B'} = \begin{bmatrix} -\frac{7}{2} \\ \frac{23}{2} \\ 6 \end{bmatrix}$

5. b. $\begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}$ **c.** $\begin{bmatrix} \frac{1}{2} & 0 \\ -\frac{1}{6} & \frac{1}{3} \end{bmatrix}$ **d.** $[\mathbf{h}]_B = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$; $[\mathbf{h}]_{B'} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$

7. a. $\begin{bmatrix} 3 & 5 \\ -1 & -2 \end{bmatrix}$ **b.** $\begin{bmatrix} 2 & 5 \\ -1 & -3 \end{bmatrix}$ **d.** $[\mathbf{w}]_{B_1} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$; $[\mathbf{w}]_{B_2} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ **e.** $[\mathbf{w}]_{B_2} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$; $[\mathbf{w}]_{B_1} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$

9. a. $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}$ **b.** $\begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix}$ **d.** $[\mathbf{w}]_B = \begin{bmatrix} -239 \\ 77 \\ 30 \end{bmatrix}$; $[\mathbf{w}]_S = \begin{bmatrix} 5 \\ -3 \\ 1 \end{bmatrix}$ **e.** $[\mathbf{w}]_S = \begin{bmatrix} 3 \\ -5 \\ 0 \end{bmatrix}$; $[\mathbf{w}]_B = \begin{bmatrix} -200 \\ 64 \\ 25 \end{bmatrix}$

11. a. $\begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}$ **13.** $P^{-1}Q^{-1}$

15. a. $B = \{(1, 1, 0), (1, 0, 2), (0, 2, 1)\}$ **b.** $B = \left\{\left(\frac{4}{5}, \frac{1}{5}, -\frac{2}{5}\right), \left(\frac{1}{5}, -\frac{1}{5}, \frac{2}{5}\right), \left(-\frac{2}{5}, \frac{2}{5}, \frac{1}{5}\right)\right\}$

17. $\begin{bmatrix} 2 & 3 \\ 5 & -1 \end{bmatrix}$ **19.** B must be the standard basis.

True/False 4.7

a. True **b.** True **c.** True **d.** True **e.** False **f.** False

Exercise Set 4.8 (page 273)

1. a. $1\begin{bmatrix}2\\-1\end{bmatrix}+2\begin{bmatrix}3\\4\end{bmatrix}$ **b.** $-2\begin{bmatrix}4\\3\\0\end{bmatrix}+3\begin{bmatrix}0\\6\\-1\end{bmatrix}+5\begin{bmatrix}-1\\2\\4\end{bmatrix}$

3. a. **b** is not in the column space of A. **b.** **b** is in the column space of A; $\begin{bmatrix}1\\9\\1\end{bmatrix}-3\begin{bmatrix}-1\\3\\1\end{bmatrix}+\begin{bmatrix}1\\1\\1\end{bmatrix}=\begin{bmatrix}5\\1\\-1\end{bmatrix}$

5. a. $r\begin{bmatrix}5\\0\\0\\0\end{bmatrix}+s\begin{bmatrix}-2\\1\\1\\0\end{bmatrix}+t\begin{bmatrix}0\\0\\1\\1\end{bmatrix}$ **b.** $\begin{bmatrix}3\\0\\-1\\5\end{bmatrix}+r\begin{bmatrix}5\\0\\0\\0\end{bmatrix}+s\begin{bmatrix}-2\\1\\1\\0\end{bmatrix}+t\begin{bmatrix}0\\0\\1\\1\end{bmatrix}$

7. a. $(1,0)+t(3,1)$; $t(3,1)$ **b.** $(-2,7,0)+t(-1,-1,1)$; $t(-1,-1,1)$

9. a. Basis for the null space: $\left\{\begin{bmatrix}16\\19\\1\end{bmatrix}\right\}$; basis for the row space: $\{[1\quad 0\quad -16], [0\quad 1\quad -19]\}$

b. Basis for the null space: $\left\{\begin{bmatrix}0\\1\\0\end{bmatrix},\begin{bmatrix}\frac{1}{2}\\0\\1\end{bmatrix}\right\}$; basis for the row space: $\{[1\quad 0\quad -\frac{1}{2}]\}$

11. a. Basis for the column space: $\left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}2\\1\\0\end{bmatrix}\right\}$; basis for the row space: $\{[1\quad 0\quad 2], [0\quad 0\quad 1]\}$

b. Basis for the column space: $\left\{\begin{bmatrix}1\\0\\0\\0\end{bmatrix},\begin{bmatrix}-3\\1\\0\\0\end{bmatrix}\right\}$; basis for the row space: $\{[1\quad -3\quad 0\quad 0], [0\quad 1\quad 0\quad 0]\}$

13. a. Basis for the row space: $\{[1\quad 0\quad 11\quad 0\quad 3], [0\quad 1\quad 3\quad 0\quad 0], [0\quad 0\quad 0\quad 1\quad 0]\}$;

basis for the column space: $\left\{\begin{bmatrix}1\\-2\\-1\\-3\end{bmatrix},\begin{bmatrix}-2\\5\\3\\8\end{bmatrix},\begin{bmatrix}0\\0\\1\\1\end{bmatrix}\right\}$

b. $\{[1\quad -2\quad 5\quad 0\quad 3], [-2\quad 5\quad -7\quad 0\quad -6], [-1\quad 3\quad -2\quad 1\quad -3]\}$

15. $\{(1,1,0,0), (0,0,1,1), (-2,0,2,2), (0,-3,0,3)\}$ **17.** Basis: $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$; $\mathbf{v}_3 = 2\mathbf{v}_1 - \mathbf{v}_2$; $\mathbf{v}_5 = -\mathbf{v}_1 + 3\mathbf{v}_2 + 2\mathbf{v}_4$

19. $\{[1\quad 4\quad 5\quad 6\quad 9], [3\quad -2\quad 1\quad 4\quad -1]\}$

21. Since $T_A(\mathbf{x}) = A\mathbf{x}$, we are seeking the general solution of the linear system $A\mathbf{x} = \mathbf{b}$.

a. $\mathbf{x} = t\left(-\frac{8}{3}, \frac{4}{3}, 1\right)$

b. $\mathbf{x} = \left(\frac{7}{3}, -\frac{2}{3}, 0\right) + t\left(-\frac{8}{3}, \frac{4}{3}, 1\right)$

c. $\mathbf{x} = \left(\frac{1}{3}, -\frac{2}{3}, 0\right) + t\left(-\frac{8}{3}, \frac{4}{3}, 1\right)$

23. a. $(x, y, z) = (1, 0, 0) + (-s - t, s, t)$

b. A plane passing through the point $(1, 0, 0)$ and parallel to the vectors $(-1, 1, 0)$ and $(-1, 0, 1)$

25. a. $x_1 = -\frac{2}{3}s + \frac{1}{3}t, x_2 = s, x_3 = t$ **c.** $(x_1, x_2, x_3) = (1, 0, 1) + \left(-\frac{2}{3}s + \frac{1}{3}t, s, t\right)$

27. $x_1 = \frac{1}{3} - \frac{4}{3}r - \frac{1}{3}s, x_2 = r, x_3 = s, x_4 = 1$;

general solution of the associated homogeneous system: $\left(-\frac{4}{3}r - \frac{1}{3}s, r, s, 0\right)$;

particular solution of the nonhomogeneous system: $\left(\frac{1}{3}, 0, 0, 1\right)$

29. b. $\begin{bmatrix}0&0&0\\0&1&0\\0&0&1\end{bmatrix}$ is an example of such a matrix.

31. **a.** $\begin{bmatrix} 3a & -5a \\ 3b & -5b \end{bmatrix}$ where a and b are not both zero.

b. Only the zero vector forms the null space for both A and B.
The line $3x + y = 0$ forms the null space for C.
The entire plane forms the null space for D.

True/False 4.8

a. True **b.** False **c.** False **d.** False **e.** False **f.** True **g.** True **h.** False **i.** True **j.** False

Exercise Set 4.9 (page 286)

1. **a.** $\text{rank}(A) = 1$; $\text{nullity}(A) = 3$ **b.** $\text{rank}(A) = 2$; $\text{nullity}(A) = 3$

3. **a.** $\text{rank}(A) = 3$; $\text{nullity}(A) = 0$ **c.** 3 leading variables; 0 parameters in the general solution (The solution is unique.)

5. **a.** $\text{rank}(A) = 1$; $\text{nullity}(A) = 2$ **c.** 1 leading variable; 2 parameters in the general solution

7. **a.** largest possible value for the rank: 4; smallest possible value for the nullity: 0

b. largest possible value for the rank: 3; smallest possible value for the nullity: 2

c. largest possible value for the rank: 3; smallest possible value for the nullity: 0

9.

		(a)	(b)	(c)	(d)	(e)	(f)	(g)
(i)	dimension of the row space of A	3	2	1	2	2	0	2
	dimension of the column space of A	3	2	1	2	2	0	2
	dimension of the null space of A	0	1	2	7	7	4	0
	dimension of the null space of A^T	0	1	2	3	3	4	4
(ii)	Is the system $A\mathbf{x} = \mathbf{b}$ consistent?	Yes	No	Yes	Yes	No	Yes	Yes
(iii)	number of parameters in the general solution of $A\mathbf{x} = \mathbf{b}$	0	—	2	7	—	4	0

11. $\dim[\text{row}(A)] = \dim[\text{col}(A)] = 2,\ \dim[\text{null}(A)] = 0,\ \dim[\text{null}(A^T)] = 1$

Basis for $\text{row}(A)$: $\{[1\ \ 0],\ [0\ \ 1]\}$

Basis for $\text{col}(A)$: $\left\{\begin{bmatrix} 1 \\ 0 \\ -9 \end{bmatrix}, \begin{bmatrix} 4 \\ 3 \\ 0 \end{bmatrix}\right\}$

Basis for $\text{null}(A)$: $\emptyset$

Basis for $\text{null}[A^T]$: $\left\{\begin{bmatrix} 9 \\ -12 \\ 1 \end{bmatrix}\right\}$

13. $\dim[\text{row}(A)] = \dim[\text{col}(A)] = 2,\ \dim[\text{null}(A)] = 1,\ \dim[\text{null}(A^T)] = 1$

Basis for $\text{row}(A)$: $\left\{\begin{bmatrix} 1 & 0 & 4 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 4 \end{bmatrix}\right\}$

Basis for $\text{col}(A)$: $\left\{\begin{bmatrix} 0 \\ -1 \\ -2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix}\right\}$

Basis for $\text{null}(A)$: $\left\{\begin{bmatrix} -4 \\ -4 \\ 1 \end{bmatrix}\right\}$

Basis for $\text{null}(A^T)$: $\left\{\begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix}\right\}$

19. Basis for $\text{row}(A)$: $\left\{\begin{bmatrix} 1 & 0 & 6 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 4 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}\right\}$

Basis for $\text{col}(A)$: $\left\{\begin{bmatrix} 0 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 4 \end{bmatrix}, \begin{bmatrix} -7 \\ 0 \\ 5 \end{bmatrix}\right\}$

Basis for $\text{null}(A)$: $\left\{\begin{bmatrix} -6 \\ -4 \\ 1 \\ 0 \end{bmatrix}\right\}$

Basis for $\text{null}(A^T)$: $\emptyset$

21. **a.** $\text{nullity}(A) - \text{nullity}(A^T) = 1$ **b.** $\text{nullity}(A) - \text{nullity}(A^T) = n - m$

23. **a.** 3 **b.** 2

25. The matrix cannot have rank 1. It has rank 2 if $r = 2$ and $s = 1$.

27. No, both row and column spaces of A must be planes.

29. **a.** 3 **b.** 5 **c.** 3 **d.** 3 **31.** **a.** 3 **b.** No

37. **a.** Overdetermined; inconsistent if $3b_1 + b_2 + 2b_3 \neq 0$

b. Underdetermined; infinitely many solutions for all b's; (cannot be inconsistent)

c. Underdetermined; infinitely many solutions for all b's; (cannot be inconsistent)

True/False 4.9

a. False **b.** True **c.** False **d.** False **e.** True **f.** False **g.** False **h.** False **i.** True **j.** False

Chapter 4 Supplementary Exercises (page 289)

1. **a.** $\mathbf{u} + \mathbf{v} = (4, 3, 2)$; $k\mathbf{u} = (-3, 0, 0)$ **c.** Axioms 1–5 **3.** a plane if $s = 1$; a line if $s = -2$; the origin if $s \neq -2$ and $s \neq 1$

7. A must be invertible.

9. **a.** Rank is 2; nullity is 1. **b.** Rank is 2; nullity is 2.

c. For $n = 1$, rank is 1 and nullity is 0; for $n \geq 2$, rank is 2 and nullity is $n - 2$.

11. **a.** $\{1, x^2, x^4, \ldots, x^{2\lfloor n/2 \rfloor}\}$ **b.** $\{1, x - x^n, x^2 - x^n, \ldots, x^{n-1} - x^n\}$

13. **a.** $\left\{\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix}, \begin{bmatrix}0&1&0\\1&0&0\\0&0&0\end{bmatrix}, \begin{bmatrix}0&0&1\\0&0&0\\1&0&0\end{bmatrix}, \begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix}, \begin{bmatrix}0&0&0\\0&0&1\\0&1&0\end{bmatrix}, \begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}\right\}$

b. $\left\{\begin{bmatrix}0&1&0\\-1&0&0\\0&0&0\end{bmatrix}, \begin{bmatrix}0&0&1\\0&0&0\\-1&0&0\end{bmatrix}, \begin{bmatrix}0&0&0\\0&0&1\\0&-1&0\end{bmatrix}\right\}$

15. Possible ranks are 0, 1, and 2. **17.** **a.** Yes **b.** No **c.** Yes

Chapter 5

Exercise Set 5.1 (page 299)

1. eigenvalue: -1 **3.** eigenvalue: 5

5. **a.** Characteristic equation: $(\lambda - 5)(\lambda + 1) = 0$;
eigenvalue: 5, basis for eigenspace: $\{(1, 1)\}$;
eigenvalue: -1, basis for eigenspace: $\{(-2, 1)\}$

b. Characteristic equation: $\lambda^2 + 3 = 0$; no real eigenvalues

c. Characteristic equation: $(\lambda - 1)^2 = 0$;
eigenvalue: 1, basis for eigenspace: $\{(1, 0), (0, 1)\}$

d. Characteristic equation: $(\lambda - 1)^2 = 0$;
eigenvalue: $\lambda = 1$, basis for eigenspace: $\{(1, 0)\}$

7. Characteristic equation: $(\lambda - 1)(\lambda - 2)(\lambda - 3) = 0$;
eigenvalue: 1, basis for eigenspace: $\{(0, 1, 0)\}$;
eigenvalue: 2, basis for eigenspace: $\{(-1, 2, 2)\}$;
eigenvalue: 3, basis for eigenspace: $\{(-1, 1, 1)\}$

9. Characteristic equation: $(\lambda + 2)^2(\lambda - 5) = 0$;
eigenvalue: -2, basis for eigenspace: $\{(1, 0, 1)\}$;
eigenvalue: 5, basis for eigenspace: $\{(8, 0, 1)\}$

11. Characteristic equation: $(\lambda - 3)^3 = 0$;
eigenvalue: 3, basis for eigenspace: $\{(0, 1, 0), (1, 0, 1)\}$

13. $(\lambda - 3)(\lambda - 7)(\lambda - 1) = 0$

15. Eigenvalue: 5, basis for eigenspace: $\{(1, 1)\}$;
eigenvalue: -1, basis for eigenspace: $\{(-2, 1)\}$

17. **b.** $\lambda = -\omega$ is the eigenvalue associated with given eigenvectors.

19. **a.** Eigenvalue: 1, eigenspace: span$\{(1, 1)\}$;
eigenvalue: -1, eigenspace: span$\{(-1, 1)\}$

b. Eigenvalue: 1, eigenspace: span$\{(1, 0)\}$;
eigenvalue: 0, eigenspace: span$\{(0, 1)\}$

c. No real eigenvalues

d. Eigenvalue: k, eigenspace: R^2
e. Eigenvalue: 1, eigenspace: span$\{(1,0)\}$

21. a. Eigenvalue: 1, eigenspace: span$\{(1,0,0),(0,1,0)\}$;
eigenvalue: -1, eigenspace: span$\{(0,0,1)\}$
b. Eigenvalue: 1, eigenspace: span$\{(1,0,0),(0,0,1)\}$;
eigenvalue: 0, eigenspace: span$\{(0,1,0)\}$
c. Eigenvalue: 1, eigenspace: span$\{(1,0,0)\}$
d. Eigenvalue: k, eigenspace: R^3

23. a. $y=2x$ and $y=x$ **b.** No invariant lines **25. a.** 6×6 **b.** Yes **c.** Three

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

True/False 5.1

a. False **b.** False **c.** True **d.** False **e.** False **f.** False

Exercise Set 5.2 (page 309)

5. $P=\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$ (answer is not unique) **7.** $P=\begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (answer is not unique)

9. a. 3 and 5 **b.** rank$(3I-A)=1$; rank$(5I-A)=2$ **c.** Yes

11. eigenvalues: 1, 2 and 3; each has algebraic multiplicity 1 and geometric multiplicity 1;
A is diagonalizable; $P=\begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 3 \\ 1 & 3 & 4 \end{bmatrix}$ (answer is not unique); $P^{-1}AP=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

13. eigenvalue $\lambda=0$ has both algebraic and geometric multiplicity 2;
eigenvalue $\lambda=1$ has both algebraic and geometric multiplicity 1;
A is diagonalizable; $P=\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 3 & 1 \end{bmatrix}$ (answer is not unique); $P^{-1}AP=\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

15. a. A is a 3×3 matrix;
all three eigenspaces (for $\lambda=1$, $\lambda=-3$, and $\lambda=5$) must have dimension 1.
b. A is a 6×6 matrix;
the possible dimensions of the eigenspace corresponding to $\lambda=0$ are 1 or 2;
the dimension of the eigenspace corresponding to $\lambda=1$ must be 1;
the possible dimensions of the eigenspace corresponding to $\lambda=2$ are 1, 2, or 3.

17. $\begin{bmatrix} 24{,}234 & -34{,}815 \\ -23{,}210 & 35{,}839 \end{bmatrix}$ **19.** $A^{11}=\begin{bmatrix} -1 & 10{,}237 & -2{,}047 \\ 0 & 1 & 0 \\ 0 & 10{,}245 & -2{,}048 \end{bmatrix}$ **21.** $\begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3^n & 0 \\ 0 & 0 & 4^n \end{bmatrix}\begin{bmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}$

25. Yes

27. a. The dimension of the eigenspace corresponding to $\lambda=1$ must be 1; the possible dimensions of the eigenspace corresponding to $\lambda=3$ are 1 or 2; the possible dimensions of the eigenspace corresponding to $\lambda=4$ are 1, 2, or 3.
b. The dimension of the eigenspace corresponding to $\lambda=1$ must be 1; the dimension of the eigenspace corresponding to $\lambda=3$ must be 2; the dimension of the eigenspace corresponding to $\lambda=4$ must be 3.
c. This eigenvalue must be $\lambda=4$.

31. Standard matrix: $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$; diagonalizable; $P=\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$ (answer is not unique)

33. Standard matrix: $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{bmatrix}$; diagonalizable; $P=\begin{bmatrix} 0 & 0 & 3 \\ 0 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ (answer is not unique)

True/False 5.2

a. False **b.** True **c.** True **d.** False **e.** True **f.** True **g.** True **h.** True **i.** True

Exercise Set 5.3 (page 322)

1. $\bar{\mathbf{u}} = (2+i, -4i, 1-i)$; $\operatorname{Re}(\mathbf{u}) = (2, 0, 1)$; $\operatorname{Im}(\mathbf{u}) = (-1, 4, 1)$; $\|\mathbf{u}\| = \sqrt{23}$ **5.** $\mathbf{x} = (7-6i, -4-8i, 6-12i)$

7. $\overline{A} = \begin{bmatrix} 5i & 4 \\ 2+i & 1-5i \end{bmatrix}$; $\operatorname{Re}(A) = \begin{bmatrix} 0 & 4 \\ 2 & 1 \end{bmatrix}$; $\operatorname{Im}(A) = \begin{bmatrix} -5 & 0 \\ -1 & 5 \end{bmatrix}$; $\det(A) = 17 - i$; $\operatorname{tr}(A) = 1$

11. $\mathbf{u}\cdot\mathbf{v} = -1+i$; $\mathbf{u}\cdot\mathbf{w} = 18-7i$; $\mathbf{v}\cdot\mathbf{w} = 12+6i$ **13.** $-11-14i$

15. Eigenvalue: $2+i$, basis for eigenspace: $\left\{\begin{bmatrix} 2+i \\ 1 \end{bmatrix}\right\}$; eigenvalue: $2-i$, basis for eigenspace: $\left\{\begin{bmatrix} 2-i \\ 1 \end{bmatrix}\right\}$

17. Eigenvalue: $4+i$, basis for eigenspace: $\left\{\begin{bmatrix} 1+i \\ 1 \end{bmatrix}\right\}$; eigenvalue: $4-i$, basis for eigenspace: $\left\{\begin{bmatrix} 1-i \\ 1 \end{bmatrix}\right\}$

19. $|\lambda| = \sqrt{2}$; $\phi = \frac{\pi}{4}$ **21.** $|\lambda| = 2$; $\phi = -\frac{\pi}{3}$ **23.** $P = \begin{bmatrix} -2 & -1 \\ 2 & 0 \end{bmatrix}$; $C = \begin{bmatrix} 3 & -2 \\ 2 & 3 \end{bmatrix}$ **25.** $P = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix}$; $C = \begin{bmatrix} 5 & -3 \\ 3 & 5 \end{bmatrix}$

27. **a.** $k = -\frac{8}{3}i$ **b.** None

True/False 5.3

a. False **b.** True **c.** False **d.** True **e.** False **f.** False

Exercise Set 5.4 (page 327)

1. **a.** $y_1 = c_1e^{5x} - 2c_2e^{-x}$, $y_2 = c_1e^{5x} + c_2e^{-x}$
b. $y_1 = 0, y_2 = 0$

3. **a.** $y_1 = -c_2e^{2x} - c_3e^{3x}$, $y_2 = c_1e^x + 2c_2e^{2x} + c_3e^{3x}$, $y_3 = 2c_2e^{2x} + c_3e^{3x}$
b. $y_1 = e^{2x} - 2e^{3x}$, $y_2 = e^x - 2e^{2x} + 2e^{3x}$, $y_3 = -2e^{2x} + 2e^{3x}$

7. $y = c_1e^{3x} - c_2e^{-2x}$ **9.** $y = c_1e^x + c_2e^{2x} + c_3e^{3x}$

15. **b.** $\mathbf{y}' = A\mathbf{y}$ where $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ and $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 1 & 2 \end{bmatrix}$

c. The solution of the system: $y_1 = c_1e^{2x} + c_2e^x + c_3e^{-x}$, $y_2 = 2c_1e^{2x} + c_2e^x - c_3e^{-x}$, and $y_3 = 4c_1e^{2x} + c_2e^x + c_3e^{-x}$;
The solution of the differential equation: $y = c_1e^{2x} + c_2e^x + c_3e^{-x}$

True/False 5.4

a. True **b.** False **c.** True **d.** True **e.** False

Exercise Set 5.5 (page 337)

1. **a.** Stochastic **b.** Not stochastic **c.** Stochastic **d.** Not stochastic

3. $\mathbf{x}_4 = \begin{bmatrix} 0.54545 \\ 0.45455 \end{bmatrix}$ **5.** **a.** Regular **b.** Not regular **c.** Regular **7.** $\begin{bmatrix} \frac{8}{17} \\ \frac{9}{17} \end{bmatrix}$ **9.** $\begin{bmatrix} \frac{4}{11} \\ \frac{4}{11} \\ \frac{3}{11} \end{bmatrix}$

11. **a.** Probability that the system will stay in state 1 when it is in state 1
b. Probability that the system will move to state 1 when it is in state 2
c. 0.8
d. 0.85

13. **a.** $\begin{bmatrix} 0.95 & 0.55 \\ 0.05 & 0.45 \end{bmatrix}$ **b.** 0.93 **c.** 0.142 **d.** 0.63

15. **a.**

	initial state	after 1 year	after 2 years	after 3 years	after 4 years	after 5 years
city population	100,000	95,750	91,840	88,243	84,933	81,889
suburb population	25,000	29,250	33,160	36,757	40,067	43,111

b. City population will approach 46,875 and the suburbs population will approach 78,125.

17. $P = \begin{bmatrix} \frac{7}{10} & \frac{1}{10} & \frac{1}{5} \\ \frac{1}{5} & \frac{3}{10} & \frac{1}{2} \\ \frac{1}{10} & \frac{3}{5} & \frac{3}{10} \end{bmatrix}$; steady-state vector: $\begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$ **19.** For any positive integer k, $P^k\mathbf{q} = \mathbf{q}$.

True/False 5.5

a. True **b.** True **c.** True **d.** False **e.** True **f.** False **g.** True

Chapter 5 Supplementary Exercises (page 339)

1. b. A is the standard matrix of the rotation in the plane about the origin through a positive angle θ. Unless the angle is an integer multiple of π, no vector resulting from such a rotation is a scalar multiple of the original nonzero vector.

3. c. $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$ **9.** $A^2 = \begin{bmatrix} 15 & 30 \\ 5 & 10 \end{bmatrix}, A^3 = \begin{bmatrix} 75 & 150 \\ 25 & 50 \end{bmatrix}, A^4 = \begin{bmatrix} 375 & 750 \\ 125 & 250 \end{bmatrix}, A^5 = \begin{bmatrix} 1875 & 3750 \\ 625 & 1250 \end{bmatrix}$

11. 0, $\operatorname{tr}(A)$ **13.** All eigenvalues must be 0 **15.** $\begin{bmatrix} 1 & 0 & 0 \\ -1 & -\frac{1}{2} & -\frac{1}{2} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}$

17. The only possible eigenvalues are -1, 0, and 1.

19. The remaining eigenvalues are 2 and 3.

Chapter 6

Exercise Set 6.1 (page 349)

1. a. 12 **b.** -18 **c.** -9 **d.** $\sqrt{30}$ **e.** $\sqrt{11}$ **f.** $\sqrt{203}$

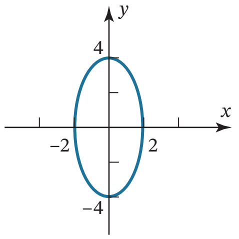
3. a. 34 **b.** -39 **c.** -18 **d.** $\sqrt{89}$ **e.** $\sqrt{34}$ **f.** $\sqrt{610}$

5. $\begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{3} \end{bmatrix}$ **7.** -24 **9.** 3 **11.** -29 **13.** $\begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{5} \end{bmatrix}$ **15.** -50 **17.** $\|\mathbf{u}\| = \sqrt{30}, d(\mathbf{u}, \mathbf{v}) = \sqrt{107}$

19. $\|\mathbf{p}\| = \sqrt{14}\,, d(\mathbf{p}, \mathbf{q}) = \sqrt{137}$ **21.** $\|U\| = \sqrt{93}, d(U, V) = \sqrt{99} = 3\sqrt{11}$ **23.** $\|\mathbf{p}\| = 6\sqrt{3}\,, d(\mathbf{p}, \mathbf{q}) = 11\sqrt{2}$

25. $\|\mathbf{u}\| = \sqrt{65}, d(\mathbf{u}, \mathbf{v}) = 12\sqrt{5}$ **27. a.** -101 **b.** 3

29.



31. $\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{9}u_1v_1 + u_2v_2$ **33.** Axioms 2 and 3 do not hold. **35.** $14\langle \mathbf{u}, \mathbf{v} \rangle - 4\,\|\mathbf{u}\|^2 - 6\,\|\mathbf{v}\|^2$

37. a. $\frac{2}{3}$ **b.** $\frac{4}{\sqrt{15}}$ **c.** $\sqrt{2}$ **d.** $\sqrt{\frac{2}{5}}$ **39.** 0 **43. b.** k_1 and k_2 must both be positive.

True/False 6.1

a. True **b.** False **c.** True **d.** True **e.** False **f.** True **g.** False

Exercise Set 6.2 (page 358)

1. a. $-\frac{1}{\sqrt{2}}$ **b.** 0 **c.** $-\frac{1}{\sqrt{2}}$ **3.** 0 **5.** $\frac{19}{10\sqrt{7}}$ **7. a.** Orthogonal **b.** Not orthogonal **c.** Orthogonal

13. Orthogonal if $k = \frac{4}{3}$ **15.** The weights must be positive numbers such that $w_1 = 4w_2$. **17.** No **25.** No

27. $\left\{(-1, -1, 1, 0), \left(\frac{2}{7}, -\frac{4}{7}, 0, 1\right)\right\}$ **29. a.** $y = -\frac{1}{2}x$ **b.** $x = t, y = -2t, z = -3t$

31. a. $\frac{1}{4}$ **b.** $\|\mathbf{p}\| = \frac{1}{\sqrt{3}}$, $\|\mathbf{q}\| = \frac{1}{\sqrt{5}}$ **33. a.** 0 **b.** $\|\mathbf{p}\| = \frac{4}{\sqrt{15}}$, $\|\mathbf{q}\| = 2\sqrt{\frac{2}{3}}$ **51. a.** $\mathbf{v} = a(1, -1)$ **b.** $\mathbf{v} = a(1, -2)$

True/False 6.2

a. False **b.** True **c.** True **d.** True **e.** False **f.** False

Exercise Set 6.3 (page 374)

1. a. Orthogonal but not orthonormal **b.** Orthogonal and orthonormal
c. Not orthogonal and not orthonormal **d.** Orthogonal but not orthonormal

3. **a.** Orthogonal **b.** Not orthogonal

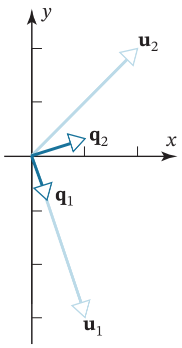
5. An orthonormal basis: $\left\{\left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right), (0, 1, 0)\right\}$

7. $\mathbf{u} = -\frac{11}{5}\mathbf{v}_1 - \frac{2}{5}\mathbf{v}_2 + 2\mathbf{v}_3$ **9.** $\mathbf{u} = 0\mathbf{v}_1 - \frac{2}{3}\mathbf{v}_2 + \frac{1}{3}\mathbf{v}_3$ **11.** $\left(-\frac{11}{5}, -\frac{2}{5}, 2\right)$ **13.** $\left(0, -\frac{2}{3}, \frac{1}{3}\right)$

15. **a.** $\left(\frac{63}{25}, \frac{84}{25}\right)$ **b.** $\left(-\frac{88}{25}, \frac{66}{25}\right)$ **17.** **a.** $\left(\frac{5}{2}, \frac{5}{2}\right)$ **b.** $\left(-\frac{1}{2}, \frac{1}{2}\right)$ **19.** **a.** $\left(\frac{10}{3}, \frac{8}{3}, \frac{4}{3}\right)$ **b.** $\left(\frac{2}{3}, -\frac{2}{3}, -\frac{1}{3}\right)$

21. **a.** $\left(\frac{22}{15}, -\frac{14}{15}, \frac{2}{3}\right)$ **b.** $\left(-\frac{7}{15}, \frac{14}{15}, \frac{7}{3}\right)$ **23.** $\left(\frac{3}{2}, \frac{3}{2}, -1, -1\right)$ **25.** $\left(\frac{23}{18}, \frac{11}{6}, -\frac{1}{18}, -\frac{17}{18}\right)$

27. $\mathbf{q}_1 = \left(\frac{1}{\sqrt{10}}, -\frac{3}{\sqrt{10}}\right), \mathbf{q}_2 = \left(\frac{3}{\sqrt{10}}, \frac{1}{\sqrt{10}}\right)$ **29.** $\left\{\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right)\right\}$



31. $\left\{\left(0, \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}, 0\right), \left(\frac{5}{\sqrt{30}}, -\frac{1}{\sqrt{30}}, \frac{2}{\sqrt{30}}, 0\right), \left(\frac{1}{\sqrt{10}}, \frac{1}{\sqrt{10}}, -\frac{2}{\sqrt{10}}, -\frac{2}{\sqrt{10}}\right), \left(\frac{1}{\sqrt{15}}, \frac{1}{\sqrt{15}}, -\frac{2}{\sqrt{15}}, \frac{3}{\sqrt{15}}\right)\right\}$

33. From Exercise 23, $\mathbf{w}_1 = \text{proj}_W\mathbf{b} = \left(\frac{3}{2}, \frac{3}{2}, -1, -1\right)$, so $\mathbf{w}_2 = \mathbf{b} - \text{proj}_W\mathbf{b} = \left(-\frac{1}{2}, \frac{1}{2}, 1, -1\right)$.

35. $\mathbf{w}_1 = \left(\frac{13}{14}, \frac{31}{14}, \frac{20}{7}\right), \mathbf{w}_2 = \left(\frac{1}{14}, -\frac{3}{14}, \frac{1}{7}\right)$ **37.** An orthonormal basis: $\left\{\left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right), \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right), \left(\frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, 0\right)\right\}$

39. For example, $\mathbf{x} = \left(\frac{1}{\sqrt{3}}, 0\right)$ and $\mathbf{y} = \left(0, \frac{1}{\sqrt{2}}\right)$ **41.** **b.** $\text{proj}_W\mathbf{u} = (2, 1, 2)$ (using both methods)

43. An orthonormal basis: $\{1, \sqrt{3}(-1 + 2x), \sqrt{5}(1 - 6x + 6x^2)\}$

45. $R = \begin{bmatrix} \sqrt{5} & \sqrt{5} \\ 0 & \sqrt{5} \end{bmatrix}$ (Q is given.) **47.** $R = \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{2} \\ 0 & \sqrt{3} & -\frac{1}{\sqrt{3}} \\ 0 & 0 & \frac{4}{\sqrt{6}} \end{bmatrix}$ (Q is given.)

49. A does not have a QR-decomposition. **55.** **b.** The range of T is W; the kernel of T is $W^\perp$.

True/False 6.3

a. False **b.** False **c.** True **d.** True **e.** False **f.** True

Exercise Set 6.4 (page 384)

1. $\begin{bmatrix} 21 & 25 \\ 25 & 35 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 20 \\ 20 \end{bmatrix}$ **3.** $x_1 = \frac{20}{11}, x_2 = -\frac{8}{11}$ **5.** $x_1 = 12, x_2 = -3, x_3 = 9$

7. Least squares error vector: $\begin{bmatrix} -\frac{6}{11} \\ -\frac{27}{11} \\ \frac{15}{11} \end{bmatrix}$; least squares error: $\frac{3}{11}\sqrt{110} \approx 2.86$

9. Least squares error vector: $\begin{bmatrix} 3 \\ -3 \\ 0 \\ 3 \end{bmatrix}$; least squares error: $3\sqrt{3} \approx 5.196$

11. Least squares solutions: $x_1 = \frac{1}{2} - \frac{1}{2}t, x_2 = t$; error vector: $\begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}$

13. Least squares solutions: $x_1 = -\frac{7}{6} - t, x_2 = \frac{7}{6} - t, x_3 = t$; error vector: $\begin{bmatrix} \frac{7}{3} \\ \frac{7}{6} \\ -\frac{49}{6} \end{bmatrix}$

15. $\begin{bmatrix} -\frac{92}{285} \\ \frac{439}{285} \\ \frac{94}{57} \end{bmatrix}$ **17.** $\begin{bmatrix} 3 \\ -4 \\ -1 \end{bmatrix}$ **19.** $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ **21.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **23.** $\begin{bmatrix} \frac{1}{7} \\ \frac{18}{7} \end{bmatrix}$

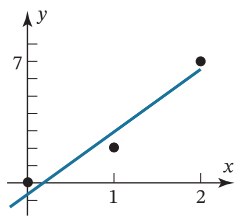
25. a. $\{(1, 0, -5), (0, 1, 3)\}$ **b.** $\frac{1}{35}\begin{bmatrix} 10 & 15 & -5 \\ 15 & 26 & 3 \\ -5 & 3 & 34 \end{bmatrix}$ **27.** $\begin{bmatrix} 0 \\ -1 \\ 1 \\ 1 \end{bmatrix}$ **29.** $A^T(AA^T)^{-1}A$

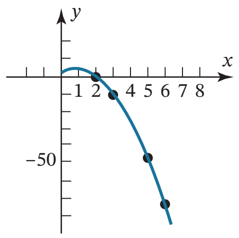
True/False 6.4

a. True **b.** False **c.** True **d.** True **e.** False **f.** True **g.** False **h.** True

Exercise Set 6.5 (page 390)

1. $y = -\frac{1}{2} + \frac{7}{2}x$ **3.** $y = 2 + 5x - 3x^2$ **5.** $y = \frac{5}{21} + \frac{48}{7x}$





True/False 6.5

a. False **b.** True **c.** True **d.** False

Exercise Set 6.6 (page 397)

1. a. $1 + \pi - 2\sin x - \sin 2x$ **b.** $1 + \pi - \frac{2}{1}\sin x - \frac{2}{2}\sin(2x) - \cdots - \frac{2}{n}\sin(nx)$

3. a. $\frac{e^x}{e-1} - \frac{1}{2}$ **b.** $\frac{7e - 19}{12e - 12} \approx 0.00136$

5. a. $\frac{3x}{\pi}$ **b.** $1 - \frac{6}{\pi^2} \approx 0.392$ **9.** $\frac{1}{2} + \sum_{k=1}^{\infty} \frac{1}{k\pi}(1 - (-1)^k)\sin kx$

True/False 6.6

a. False **b.** True **c.** True **d.** False **e.** True

Chapter 6 Supplementary Exercises (page 397)

1. a. $(0, a, a, 0)$ with $a \neq 0$ **b.** $\pm\left(0, \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}, 0\right)$

3. a. The subspace of all matrices in M_{22} with zeros on the main diagonal.
b. The subspace of all 2×2 skew-symmetric matrices.

7. $\pm\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$ **9.** No **11. b.** θ approaches $\frac{\pi}{2}$ **17.** No

Chapter 7

Exercise Set 7.1 (page 405)

1. a. Orthogonal; $A^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ **b.** Orthogonal; $A^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

3. a. Not orthogonal **b.** Orthogonal; $A^{-1} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$ **7.** $T_A(\mathbf{x}) = \begin{bmatrix} -\frac{23}{5} \\ \frac{18}{25} \\ \frac{101}{25} \end{bmatrix}$; $\|T_A(\mathbf{x})\| = \|\mathbf{x}\| = \sqrt{38}$

9. Yes **11.** $a^2 + b^2 = \frac{1}{2}$ **13. a.** $\begin{bmatrix} -1+3\sqrt{3} \\ 3+\sqrt{3} \end{bmatrix}$ **b.** $\begin{bmatrix} \frac{5}{2}-\sqrt{3} \\ 1+\frac{5}{2}\sqrt{3} \end{bmatrix}$ **15. a.** $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{3}{\sqrt{2}} \\ 5 \end{bmatrix}$ **b.** $\begin{bmatrix} -\frac{5}{\sqrt{2}} \\ \frac{7}{\sqrt{2}} \\ -3 \end{bmatrix}$

17. a. $\begin{bmatrix} -\frac{1}{2}-\frac{5\sqrt{3}}{2} \\ 2 \\ -\frac{\sqrt{3}}{2}+\frac{5}{2} \end{bmatrix}$ **b.** $\begin{bmatrix} \frac{1}{2}-\frac{3\sqrt{3}}{2} \\ 6 \\ -\frac{\sqrt{3}}{2}-\frac{3}{2} \end{bmatrix}$ **19.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}$

21. a. Rotations about the origin, reflections about any line through the origin, and any combination of these
b. Rotations about the origin, dilations, contractions, reflections about lines through the origin, and combinations of these
c. No; dilations and contractions

23. a. $(\mathbf{p})_S = \left(\frac{5}{\sqrt{3}}, \sqrt{2}, \frac{\sqrt{2}}{\sqrt{3}}\right), (\mathbf{q})_S = \left(-\frac{2}{\sqrt{3}}, 2\sqrt{2}, -\frac{\sqrt{2}}{\sqrt{3}}\right)$
b. $\|\mathbf{p}\| = \sqrt{11}, d(\mathbf{p},\mathbf{q}) = \sqrt{21}, \langle \mathbf{p},\mathbf{q}\rangle = 0$

True/False 7.1

a. False **b.** False **c.** False **d.** False **e.** True **f.** True **g.** True **h.** True

Exercise Set 7.2 (page 415)

1. $\lambda^2 - 5\lambda = 0$; $\lambda = 0$: one-dimensional; $\lambda = 5$: one-dimensional

3. $\lambda^3 - 3\lambda^2 = 0$; $\lambda = 3$: one-dimensional; $\lambda = 0$: two-dimensional

5. $\lambda^4 - 8\lambda^3 = 0$; $\lambda = 0$: three-dimensional; $\lambda = 8$: one-dimensional

7. $P = \begin{bmatrix} -\frac{2}{\sqrt{7}} & \frac{\sqrt{3}}{\sqrt{7}} \\ \frac{\sqrt{3}}{\sqrt{7}} & \frac{2}{\sqrt{7}} \end{bmatrix}$; $P^{-1}AP = \begin{bmatrix} 3 & 0 \\ 0 & 10 \end{bmatrix}$ **9.** $P = \begin{bmatrix} -\frac{4}{5} & 0 & \frac{3}{5} \\ 0 & 1 & 0 \\ \frac{3}{5} & 0 & \frac{4}{5} \end{bmatrix}$; $P^{-1}AP = \begin{bmatrix} 25 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -50 \end{bmatrix}$

11. $P = \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$; $P^{-1}AP = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

13. $P = \begin{bmatrix} -\frac{4}{5} & 0 & \frac{3}{5} & 0 \\ \frac{3}{5} & 0 & \frac{4}{5} & 0 \\ 0 & -\frac{4}{5} & 0 & \frac{3}{5} \\ 0 & \frac{3}{5} & 0 & \frac{4}{5} \end{bmatrix}$; $P^{-1}AP = \begin{bmatrix} -25 & 0 & 0 & 0 \\ 0 & -25 & 0 & 0 \\ 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 25 \end{bmatrix}$

15. $(2)\begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} + (4)\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = (2)\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} + (4)\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

17. $(-4)\begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}\begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix} + (-4)\begin{bmatrix} -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}\begin{bmatrix} -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} + (2)\begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \end{bmatrix}$

$= (-4)\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} + (-4)\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{bmatrix} + (2)\begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$

19. $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 4 \\ 0 & 4 & 3 \end{bmatrix}$ **21.** Yes **23. a.** $\begin{bmatrix} \frac{\sqrt{2}-1}{4-2\sqrt{2}} \\ \frac{1}{4-2\sqrt{2}} \end{bmatrix}, \begin{bmatrix} \frac{-\sqrt{2}-1}{4+2\sqrt{2}} \\ \frac{1}{4+2\sqrt{2}} \end{bmatrix}$ **b.** $\begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

True/False 7.2

a. True **b.** True **c.** False **d.** True **e.** True **f.** True **g.** True

Exercise Set 7.3 (page 427)

1. a. $[x_1 \quad x_2]\begin{bmatrix}3 & 0\\0 & 7\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}$ **b.** $[x_1 \quad x_2]\begin{bmatrix}4 & -3\\-3 & -9\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}$ **c.** $[x_1 \quad x_2 \quad x_3]\begin{bmatrix}9 & 3 & -4\\3 & -1 & \frac{1}{2}\\-4 & \frac{1}{2} & 4\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$

3. $2x^2+5y^2-6xy$ **5.** $\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}y_1\\y_2\end{bmatrix}$; $Q=3y_1^2+y_2^2$

7. $\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}-\frac{2}{3} & \frac{2}{3} & -\frac{1}{3}\\\frac{2}{3} & \frac{1}{3} & -\frac{2}{3}\\\frac{1}{3} & \frac{2}{3} & \frac{2}{3}\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}$; $Q=y_1^2+4y_2^2+7y_3^2$

9. a. $[x \quad y]\begin{bmatrix}2 & \frac{1}{2}\\\frac{1}{2} & 0\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}+[1 \quad -6]\begin{bmatrix}x\\y\end{bmatrix}+(2)=0$ **b.** $[x \quad y]\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}+[7 \quad -8]\begin{bmatrix}x\\y\end{bmatrix}+(-5)=0$

11. a. Ellipse **b.** Hyperbola **c.** Parabola **d.** Circle **13.** Hyperbola: $3y'^2-2x'^2=8$; $\theta=\sin^{-1}\left(\frac{2}{\sqrt{5}}\right)\approx 63.4^\circ$

15. Hyperbola: $4x'^2-y'^2=3$; $\theta=\sin^{-1}\left(\frac{3}{5}\right)\approx 36.9^\circ$

17. a. Positive definite **b.** Negative definite **c.** Indefinite **d.** Positive semidefinite **e.** Negative semidefinite

19. Positive definite **21.** Positive semidefinite **23.** Indefinite **27. a.** Indefinite **b.** Negative definite **29.** $k>2$

33. a. $s_x^2=\mathbf{x}^T\begin{bmatrix}\frac{1}{n} & -\frac{1}{n(n-1)} & \cdots & -\frac{1}{n(n-1)}\\-\frac{1}{n(n-1)} & \frac{1}{n} & \cdots & -\frac{1}{n(n-1)}\\\vdots & \vdots & \ddots & \vdots\\-\frac{1}{n(n-1)} & -\frac{1}{n(n-1)} & \cdots & \frac{1}{n}\end{bmatrix}\mathbf{x}$ **35.** A must have a positive eigenvalue of multiplicity 2.

True/False 7.3

a. True **b.** False **c.** True **d.** True **e.** False **f.** True **g.** True **h.** True **i.** True **j.** True **k.** True
l. False

Exercise Set 7.4 (page 435)

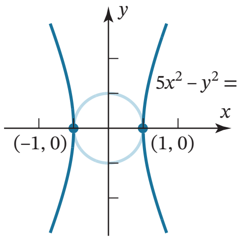
1. Maximum: 5 at $(x,y)=(\pm1,0)$; minimum: -1 at $(x,y)=(0,\pm1)$

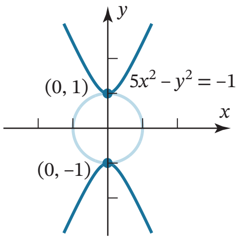
3. Maximum: 7 at $(x,y)=(0,\pm1)$; minimum: 3 at $(x,y)=(\pm1,0)$

5. Maximum: 9 at $(x,y,z)=(\pm1,0,0)$; minimum: 3 at $(x,y,z)=(0,0,\pm1)$

7. Maximum: $\sqrt{2}$ at $(x,y)=(\sqrt{2},1)$ and $(x,y)=(-\sqrt{2},-1)$; minimum: $-\sqrt{2}$ at $(x,y)=(-\sqrt{2},1)$ and $(x,y)=(\sqrt{2},-1)$

9.





13. Saddle point at $(0,0)$; relative maximum at $(-1,1)$

15. Relative minimum at $(0,0)$; saddle point at $(2,1)$; saddle point at $(-2,1)$ **17.** $x=\frac{5}{\sqrt{2}}, y=\frac{1}{\sqrt{2}}$ **21.** $q(\mathbf{x})=\lambda$

True/False 7.4

a. False **b.** True **c.** True **d.** False **e.** True

Exercise Set 7.5 (page 442)

1. $\begin{bmatrix}-2i & 4 & 5-i\\1+i & 3-i & 0\end{bmatrix}$ **3.** $\begin{bmatrix}1 & i & 2-3i\\-i & -3 & 1\\2+3i & 1 & 2\end{bmatrix}$ **5. a.** $(A)_{13}\neq(A^*)_{13}$ **b.** $(A)_{22}\neq(A^*)_{22}$

9. $A^{-1} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ -\frac{4}{5}i & -\frac{3}{5}i \end{bmatrix}$ **11.** $A^{-1} = \begin{bmatrix} \frac{1}{2\sqrt{2}}(\sqrt{3}-i) & \frac{1}{2\sqrt{2}}(1-i\sqrt{3}) \\ \frac{1}{2\sqrt{2}}(1+i\sqrt{3}) & \frac{1}{2\sqrt{2}}(-i-\sqrt{3}) \end{bmatrix}$ **13.** $P = \begin{bmatrix} \frac{-1+i}{\sqrt{3}} & \frac{1-i}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{bmatrix}; P^{-1}AP = \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix}$

15. $P = \begin{bmatrix} \frac{-1-i}{\sqrt{6}} & \frac{1+i}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}; P^{-1}AP = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$ **17.** $P = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1-i}{\sqrt{3}} & \frac{-1+i}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \end{bmatrix}; P^{-1}AP = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

19. $\begin{bmatrix} 0 & i & 2-3i \\ i & 0 & 1 \\ -2-3i & -1 & 4i \end{bmatrix}$ **27.** **c.** B and C must commute. **35.** $\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$

True/False 7.5

a. False **b.** False **c.** True **d.** False **e.** False

Chapter 7 Supplementary Exercises (page 444)

1. **a.** $\begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix}$ **b.** $\begin{bmatrix} \frac{4}{5} & 0 & -\frac{3}{5} \\ -\frac{9}{25} & \frac{4}{5} & -\frac{12}{25} \\ \frac{12}{25} & \frac{3}{5} & \frac{16}{25} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{4}{5} & -\frac{9}{25} & \frac{12}{25} \\ 0 & \frac{4}{5} & \frac{3}{5} \\ -\frac{3}{5} & -\frac{12}{25} & -\frac{16}{25} \end{bmatrix}$

5. $P = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}; P^TAP = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

7. Positive definite **9.** **a.** Parabola **b.** Parabola

13. Two possible solutions: $a = 0, b = \sqrt{\frac{2}{3}}, c = -\frac{1}{\sqrt{3}}$ and $a = 0, b = -\sqrt{\frac{2}{3}}, c = \frac{1}{\sqrt{3}}$

Chapter 8

Exercise Set 8.1 (page 456)

1. **a.** Nonlinear

b. Linear; kernel consists of all matrices of the form $\begin{bmatrix} a & b \\ c & -a \end{bmatrix}$

c. Linear; kernel consists of all matrices of the form $\begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix}$

3. Nonlinear **5.** Linear; kernel consists of all 2×2 matrices whose rows are orthogonal to all columns of B

7. **a.** Linear; $\ker(T) = \{0\}$ **b.** Nonlinear **9.** Linear; $\ker(T) = \{(0, 0, 0, \dots)\}$ **11.** (a) and (d)

13. **a.** 2 **b.** 4 **c.** $mn - 3$ **d.** 1 **15.** **a.** $\begin{bmatrix} 3 & 6 \\ -12 & 9 \end{bmatrix}$ **b.** $\text{rank}(T) = 4$; $\text{nullity}(T) = 0$

17. **a.** $(1, 0, 1)$ **b.** $\ker(T) = \{0\}$ **c.** $R(T) = R^3$ **19.** $T(x_1, x_2) = (-4x_1 + 5x_2, x_1 - 3x_2)$; $T(5, -3) = (-35, 14)$

21. $T(x_1, x_2, x_3) = (-x_1 + 4x_2 - x_3, 5x_1 - 5x_2 - x_3, x_1 + 3x_3)$; $T(2, 4, -1) = (15, -9, -1)$

23. **a.** $\left\{\begin{bmatrix} 1 \\ 5 \\ 7 \end{bmatrix}, \begin{bmatrix} -1 \\ 6 \\ 4 \end{bmatrix}\right\}$ **b.** $\{(-14, 19, 11)\}$ **c.** $\text{rank}(T) = 2$; $\text{nullity}(T) = 1$ **d.** $\text{rank}(A) = 2$; $\text{nullity}(A) = 1$

25. Basis for $\ker(T_A)$: $\{(10, 2, 0, 7)\}$; basis for $R(T_A)$: $\left\{\begin{bmatrix} 1 \\ -3 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 8 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ 4 \end{bmatrix}\right\}$

27. **b.** $\{x, x^2\}$ **c.** $\{5, x^2\}$

29. **a.** $\ker(D)$ consists of all constant polynomials
b. $\ker(J)$ consists of all polynomials of the form a_1x

31. **a.** $T(f(x)) = f^{(4)}(x)$
b. $T(f(x)) = f^{(n+1)}(x)$

33. **a.** The origin, a line through the origin, a plane through the origin, or the entire space R^3
b. The origin, a line through the origin, a plane through the origin, or the entire space R^3

35. $(-10, -7, 6)$

True/False 8.1

a. True **b.** False **c.** True **d.** False **e.** True **f.** True **g.** False **h.** False **i.** False

Exercise Set 8.2 (page 468)

1. **a.** Not one-to-one **b.** One-to-one **c.** One-to-one **3.** **a.** $\ker(T) = \{\mathbf{0}\}$; T is one-to-one **b.** $\ker(T) = \{\mathbf{0}\}$; T is one-to-one **c.** $\ker(T) = \{\text{span}(0, 1, 1)\}$; T is not one-to-one

5. **a.** $\text{nullity}(A) = 1$; not one-to-one **b.** $\text{nullity}(A) = 1$; not one-to-one

7. **a.** One-to-one **b.** One-to-one **c.** Not one-to-one

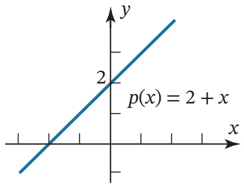
9. For example, $T(1 - x^2) = (0, 0)$; T is onto

11. No; T is not one-to-one because $\ker(T) \neq \{\mathbf{0}\}$; for example, $T(\mathbf{a}) = \mathbf{a} \times \mathbf{a} = \mathbf{0}$

13. **a.** One-to-one, not onto **b.** Not one-to-one, onto **c.** One-to-one, onto **d.** Not one-to-one, not onto

15. **a.** Reflection about the x-axis **b.** Rotation through an angle of $-\pi/4$

19. **a.** $(1, -1)$ **d.** $T^{-1}(2, 3) = 2 + x$



21. **a.** all the a_i's must be nonzero **b.** $T^{-1}(x_1, x_2, \ldots, x_n) = \left(\frac{1}{a_1}x_1, \frac{1}{a_2}x_2, \ldots, \frac{1}{a_n}x_n\right)$

23. $(T_2 \circ T_1)(x, y) = (2x - 3y, 2x + 3y)$

25. $a_0 x + a_1 x(x + 1) + a_2 x(x + 1)^2$

27. **a.** $a + d$ **b.** $(T_2 \circ T_1)(A)$ does not exist because $T_1(A)$ is not a 2×2 matrix

29. $(T_3 \circ T_2 \circ T_1)(x, y) = (3x - 2y, x)$

31. **a.** $T_1^{-1}(p(x)) = \frac{1}{x}p(x)$; $T_2^{-1}(p(x)) = p(x - 1)$; $(T_1^{-1} \circ T_2^{-1})(p(x)) = \frac{1}{x}p(x - 1)$

33. $T_2(\mathbf{v}) = \frac{1}{4}\mathbf{v}$ **39.** Since $\ker(J) \neq \{\mathbf{0}\}$, J is not one-to-one.

41. **a.** Range of T must be a proper subset of R^n **b.** T maps infinitely many vectors into $\mathbf{0}$

43. **a.** Yes **b.** Yes

True/False 8.2

a. True **b.** True **c.** True **d.** True **e.** False **f.** True **g.** True **h.** False **i.** True **j.** True

Exercise Set 8.3 (page 476)

1. Isomorphism **3.** Isomorphism **5.** Not an isomorphism **7.** Isomorphism

9. **a.** $T\left(\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}\right) = \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix}$ **b.** $T_1\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$; $T_2\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a \\ c \\ b \\ d \end{bmatrix}$

11. Isomorphism **13.** $\dim(W) = 3$; $(-r - s - t, r, s, t) \to (r, s, t)$ is an isomorphism between W and R^3

15. Isomorphism **17.** Yes **19.** No

True/False 8.3

a. False **b.** True **c.** False **d.** True **e.** True **f.** True

Exercise Set 8.4 (page 484)

1. **a.** $\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **3.** **a.** $\begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$ **5.** **a.** $\begin{bmatrix} 0 & 0 \\ -\frac{1}{2} & 1 \\ \frac{8}{3} & \frac{4}{3} \end{bmatrix}$ **7.** **a.** $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \\ 0 & 0 & 4 \end{bmatrix}$ **b, c.** $3 + 10x + 16x^2$

9. **a.** $[T(\mathbf{v}_1)]_B = \begin{bmatrix} 1 \\ -2 \end{bmatrix}; [T(\mathbf{v}_2)]_B = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$ **b.** $T(\mathbf{v}_1) = \begin{bmatrix} 3 \\ -5 \end{bmatrix}; T(\mathbf{v}_2) = \begin{bmatrix} -2 \\ 29 \end{bmatrix}$

c. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} \frac{18}{7} & \frac{1}{7} \\ -\frac{107}{7} & \frac{24}{7} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ **d.** $T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} \frac{19}{7} \\ -\frac{83}{7} \end{bmatrix}$

11. **a.** $[T(\mathbf{v}_1)]_B = \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix}; [T(\mathbf{v}_2)]_B = \begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix}; [T(\mathbf{v}_3)]_B = \begin{bmatrix} -1 \\ 5 \\ 4 \end{bmatrix}$

b. $T(\mathbf{v}_1) = 16 + 51x + 19x^2; T(\mathbf{v}_2) = -6 - 5x + 5x^2; T(\mathbf{v}_3) = 7 + 40x + 15x^2$

c. $T(a_0 + a_1x + a_2x^2) = \frac{239a_0 - 161a_1 + 289a_2}{24} + \frac{201a_0 - 111a_1 + 247a_2}{8}x + \frac{61a_0 - 31a_1 + 107a_2}{12}x^2$

d. $T(1 + x^2) = 22 + 56x + 14x^2$

13. **a.** $[T_2 \circ T_1]_{B',B} = \begin{bmatrix} 0 & 0 \\ 6 & 0 \\ 0 & -9 \\ 0 & 0 \end{bmatrix}; [T_1]_{B'',B} = \begin{bmatrix} 2 & 0 \\ 0 & -3 \\ 0 & 0 \end{bmatrix}; [T_2]_{B',B''} = \begin{bmatrix} 0 & 0 & 0 \\ 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ **b.** $[T_2 \circ T_1]_{B',B} = [T_2]_{B',B''}[T_1]_{B'',B}$

15. **a.** $[T]_{B,B'} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \end{bmatrix}; [T]_{B,B''} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix}$ **b, c.** $\begin{bmatrix} 2 & 5 \\ 1 & 2 \end{bmatrix}$

17. **a.** $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$ **b.** $-6 + 48x$ **19.** **a.** $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$ **b.** $4\sin x + 3\cos x$

21. **a.** $[T_2 \circ T_1]_{B',B} = [T_2]_{B',B''}[T_1]_{B'',B}$ **b.** $[T_3 \circ T_2 \circ T_1]_{B',B} = [T_3]_{B',B'''}[T_2]_{B''',B''}[T_1]_{B'',B}$

23. The matrix for T relative to B is the matrix whose columns are the transforms of the basis vectors in B in terms of the standard basis. Since B is the standard basis for R^n, this matrix is the standard matrix for T. Also, since B' is the standard basis for R^m, the resulting transformation will give vector components relative to the standard basis.

True/False 8.4

a. False **b.** False **c.** True **d.** False **e.** True

Exercise Set 8.5 (page 491)

1. **a.** $\det(A) = -2$ does not equal $\det(B) = -1$ **b.** $\operatorname{tr}(A) = 3$ does not equal $\operatorname{tr}(B) = -2$

3. $\begin{bmatrix} 6 & -10 \\ 2 & -3 \end{bmatrix}$ **5.** $\begin{bmatrix} -2 & -2 \\ 6 & 5 \end{bmatrix}$ **7.** $[T]_B = \begin{bmatrix} 1 & -2 \\ 0 & -1 \end{bmatrix}; [T]_{B'} = \begin{bmatrix} 11 & 20 \\ -6 & -11 \end{bmatrix}$

9. $[T]_B = \begin{bmatrix} -2 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}; [T]_{B'} = \begin{bmatrix} -2 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ **11.** $[T]_B = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}; [T]_{B'} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

13. $[T]_B = \begin{bmatrix} -1 & 0 \\ 1 & 1 \end{bmatrix}; [T]_{B'} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}$

15. **a.** $-4, 3$

b. A basis for the eigenspace corresponding to $\lambda = -4$ is $\{-2 + \frac{8}{3}x + x^2\}$;

A basis for the eigenspace corresponding to $\lambda = 3$ is $\{5 - 2x + x^2\}$

19. $\det(T) = 17$; eigenvalues: $5 \pm 2\sqrt{2}$ **21.** $\det(T) = 1$; eigenvalue: 1

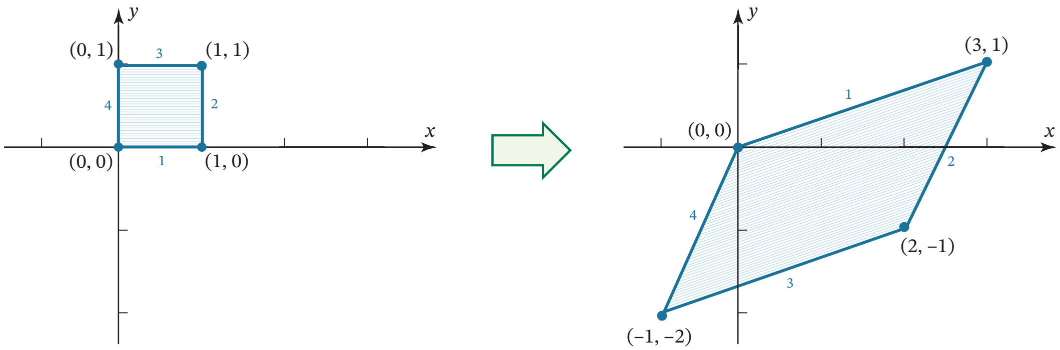
True/False 8.5

a. False **b.** True **c.** True **d.** True **e.** True **f.** False **g.** True **h.** False

Exercise Set 8.6 (page 504)

1. $y' = \frac{6}{13}x'$ **3.** $y' = \frac{2}{7}x'$

5.



7. a. $\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 5 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 0 \\ 2 & 5 \end{bmatrix}$ **c.** $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$

9. a. Operators commute **b.** Operators do not commute

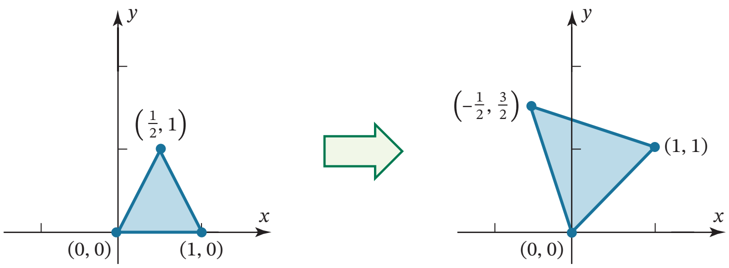
11. Shearing by a factor of 1 in the x-direction, then reflection about the x-axis, then expanding by a factor of 2 in the y-direction, then expanding by a factor of 4 in the x-direction.

13. Reflection about the x-axis, then expanding by a factor of 2 in the y-direction, then expanding by a factor of 4 in the x-direction, then reflection about the line $y = x$.

15. a. The unit square is expanded in the x-direction by a factor of 3.
b. The unit square is reflected about the x-axis and expanded in the y-direction by a factor of 5.

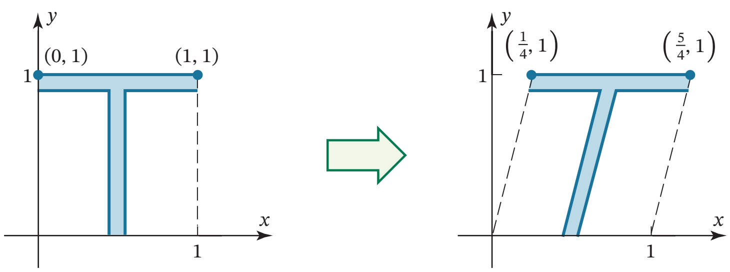
17. b. No, Theorem 8.6.1 applies only to invertible matrices.

21. a.



b. Shearing by a factor of -1 in the x-direction, then expanding by a factor of 2 in the y-direction, then shearing by a factor of 1 in the y-direction.

23.



25. The line segment from (0,0) to (2,0). Theorem 8.6.1 does not apply here because A is singular.

27. a. $\begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

29. $\begin{bmatrix} -\frac{1}{9} & \frac{8}{9} & \frac{4}{9} \\ \frac{8}{9} & -\frac{1}{9} & \frac{4}{9} \\ \frac{4}{9} & \frac{4}{9} & -\frac{7}{9} \end{bmatrix}$

True/False 8.6

a. False **b.** True **c.** True **d.** True **e.** False **f.** False **g.** True

Chapter 8 Supplementary Exercises (page 506)

1. No

5. a. $T(\mathbf{e}_3)$ and any two of $T(\mathbf{e}_1), T(\mathbf{e}_2), T(\mathbf{e}_4)$ form a basis for the range; a basis for ker(T) is $\begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$

b. rank$(T) = 3$; nullity$(T) = 1$

7. a. rank$(T) = 2$; nullity$(T) = 2$ **b.** T is not one-to-one

11. $\operatorname{rank}(T) = 3$; $\operatorname{nullity}(T) = 1$ **13.** $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ **15.** $\begin{bmatrix} -4 & 0 & 9 \\ 1 & 0 & -2 \\ 0 & 1 & 1 \end{bmatrix}$ **17.** $\begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$ **19. b.** $\{1, x\}$

25. $\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & & \cdots & \frac{1}{n} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{n+1} \end{bmatrix}$

Chapter 9

Exercise Set 9.1 (page 518)

1. $x_1 = 2, x_2 = 1$ **3.** $x_1 = 3, x_2 = -1$ **5.** $x_1 = -1, x_2 = 1, x_3 = 0$

7. a. $L^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$; $U^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{8} & -\frac{7}{48} \\ 0 & \frac{1}{4} & \frac{5}{24} \\ 0 & 0 & \frac{1}{6} \end{bmatrix}$ **b.** $A^{-1} = \begin{bmatrix} \frac{5}{48} & -\frac{1}{48} & -\frac{7}{48} \\ -\frac{7}{24} & \frac{11}{24} & \frac{5}{24} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}$

9. a. $A = LU = \begin{bmatrix} 2 & 0 & 0 \\ -2 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ **b.** $A = L_1DU_1 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

c. $A = L_2U_2 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

11. $x_1 = \frac{21}{17}, x_2 = -\frac{14}{17}, \quad x_3 = \frac{12}{17}$ **13.** $A = LDU = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

15. $A = PLU = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 3 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}$; $x_1 = -\frac{1}{2}, \quad x_2 = \frac{1}{2} \quad x_3 = 3$

17. Approximately $\frac{2}{3}n^3$ additions and multiplications are required.

True/False 9.1

a. False **b.** False **c.** True **d.** True **e.** True

Exercise Set 9.2 (page 527)

1. a. $\lambda_3 = -8$ is the dominant eigenvalue. **b.** no dominant eigenvalue

3. $\mathbf{x}_1 \approx \begin{bmatrix} 0.98058 \\ -0.19612 \end{bmatrix}, \lambda^{(1)} \approx 5.15385$; $\mathbf{x}_2 \approx \begin{bmatrix} 0.98837 \\ -0.15206 \end{bmatrix}, \lambda^{(2)} \approx 5.16185$;

$\mathbf{x}_3 \approx \begin{bmatrix} 0.98679 \\ -0.16201 \end{bmatrix}, \lambda^{(3)} \approx 5.16226$; $\mathbf{x}_4 \approx \begin{bmatrix} 0.98715 \\ -0.15977 \end{bmatrix}, \lambda^{(4)} \approx 5.16228$;

dominant eigenvalue: $2 + \sqrt{10} \approx 5.16228$;
corresponding unit eigenvector: $\frac{1}{\sqrt{20 + 6\sqrt{10}}}(3\sqrt{10}, -1) \approx (0.98709, -0.16018)$

5. $\mathbf{x}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \lambda^{(1)} = 6$; $\mathbf{x}_2 = \begin{bmatrix} -0.5 \\ 1 \end{bmatrix}, \lambda^{(2)} = 6.6$; $\mathbf{x}_3 \approx \begin{bmatrix} -0.53846 \\ 1 \end{bmatrix}, \lambda^{(3)} \approx 6.60550$;

$\mathbf{x}_4 \approx \begin{bmatrix} -0.53488 \\ 1 \end{bmatrix}, \lambda^{(4)} \approx 6.60555$;

dominant eigenvalue: $3+\sqrt{13}\approx 6.60555$;
corresponding scaled eigenvector: $\left(\frac{2-\sqrt{13}}{3}, 1\right)\approx(-0.53518, 1)$

7. **a.** $\mathbf{x}_1=\begin{bmatrix}1\\-0.5\end{bmatrix}$; $\mathbf{x}_2=\begin{bmatrix}1\\-0.8\end{bmatrix}$; $\mathbf{x}_3\approx\begin{bmatrix}1\\-0.929\end{bmatrix}$

b. $\lambda^{(1)}=2.8$; $\lambda^{(2)}\approx 2.976$; $\lambda^{(3)}\approx 2.997$

c. eigenvector: $(1,-1)$; eigenvalue: 3

d. 0.1%

9. 2.99993; $\begin{bmatrix}0.99180\\1.00000\end{bmatrix}$

13. **a.** Starting with $\mathbf{x}_0=\begin{bmatrix}1\\0\\0\end{bmatrix}$ it takes 8 iterations. **b.** Starting with $\mathbf{x}_0=\begin{bmatrix}1\\0\\0\\0\end{bmatrix}$ it takes 8 iterations.

Exercise Set 9.3 (page 532)

1. **a.** ≈ 0.067 second **b.** ≈ 66.68 seconds **c.** $\approx 66{,}668$ seconds, or about 18.5 hours

3. **a.** ≈ 9.52 seconds **b.** ≈ 0.0014 second **c.** ≈ 9.52 seconds **d.** ≈ 28.57 seconds

5. **a.** about 6.67×10^5 seconds for forward phase; about 10 seconds for backward phase

b. 1334 gigaflops per second

7. n^2 flops **9.** $2n^3-n^2$ flops

Exercise Set 9.4 (page 539)

1. $\sqrt{5}, 0$ **3.** $\sqrt{5}$ **5.** $A=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\sqrt{2} & 0\\ 0 & \sqrt{2}\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$ **7.** $A=\begin{bmatrix}\frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}}\\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}}\end{bmatrix}\begin{bmatrix}8 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}}\\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}}\end{bmatrix}$

9. $A=\begin{bmatrix}\frac{2}{3} & \frac{1}{\sqrt{2}} & \frac{\sqrt{2}}{6}\\ \frac{1}{3} & 0 & -\frac{2\sqrt{2}}{3}\\ -\frac{2}{3} & \frac{1}{\sqrt{2}} & -\frac{\sqrt{2}}{6}\end{bmatrix}\begin{bmatrix}3\sqrt{2} & 0\\ 0 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}$ **11.** $A=\begin{bmatrix}\frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}}\\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}}\end{bmatrix}\begin{bmatrix}\sqrt{3} & 0\\ 0 & \sqrt{2}\\ 0 & 0\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$

19. **b.** $A=\begin{bmatrix}\sqrt{2} & 0\\ 0 & \sqrt{2}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}$

True/False 9.4

a. False **b.** True **c.** False **d.** False **e.** True **f.** False **g.** True

Exercise Set 9.5 (page 543)

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

7. $A=\sqrt{3}\begin{bmatrix}\frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\\ -\frac{1}{\sqrt{3}}\end{bmatrix}\begin{bmatrix}1 & 0\end{bmatrix}+\sqrt{2}\begin{bmatrix}0\\ \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}0 & 1\end{bmatrix}$ **9.** 70,100 numbers must be stored; A has 100,000 entries.

True/False 9.5

a. True **b.** True **c.** False

Chapter 9 Supplementary Exercises (page 543)

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

5. a. dominant eigenvalue: 3, corresponding positive unit eigenvector: $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

b. $\mathbf{x}_5 \approx \begin{bmatrix} 0.7100 \\ 0.7042 \end{bmatrix}$; $\mathbf{v} \approx \begin{bmatrix} 0.7071 \\ 0.7071 \end{bmatrix}$

c. $\mathbf{x}_5 \approx \begin{bmatrix} 1 \\ 0.9918 \end{bmatrix}$

7. The Rayleigh quotients will slowly converge to the dominant eigenvalue $\lambda_4 = -8.1$.

9. $A = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

11. $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 24 & 0 \\ 0 & 12 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix}$

INDEX

W

Y

Z